Day 5: Data visualization and communication

Communicating and Reporting Strategies

Sebastian Ramirez-Ruiz
Hertie School

Table of contents

Let's take a brief assessment about what we have learned

Communicating data science

The final piece of the pipeline

Preparatory work

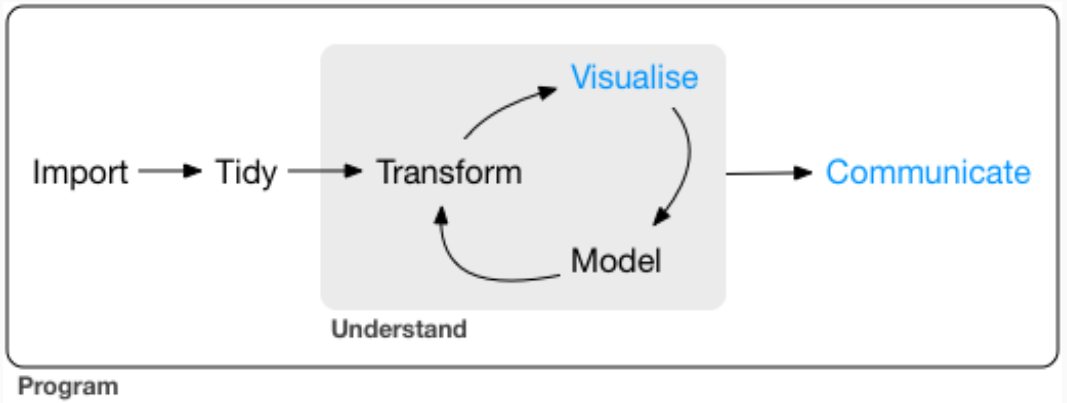
- **Problem definition** predict, infer, describe
- **Design** conceptualize, build data collection device
- **Data collection** recruit, collect, monitor

Data operation

- **Wrangle**: import, tidy, manipulate
- **Explore**: visualize, describe, discover
- **Model**: build, test, infer, predict

Dissemination

- 👉👉 **Communicate**: to the public, media, policymakers
- 👉👉 **Publish**: articles, blogs, software
- **Productize**: make usable, robust, scalable



"[I]t doesn't matter how great your analysis is unless you can explain it to others: you need to communicate your results."

Hadley Wickham & Garrett Grolemund, *R for Data Science*

Lasswell model of communication for DS

Laswell's framework of communication[1] dissects the task of communication along the following dimensions: (1) Who communicates (2) what (3) in what form (4) to whom (5) to what effect?

Let's apply this to us. Data scientists communicate...

What

- Estimates
- Uncertainty
- Model implications
- Substantive knowledge
- Product
- Themselves

How

- Spoken word
- Technical reports
- Academic papers
- Web applications
- Policy briefs

To whom

- The public
- The media
- Policymakers
- Other scientists
- Managers / co-workers

To what end

- Inform
- Influence
- Instruct
- Motivate
- Monitor
- Document

What, how, and to what end you communicate depends on your **audience/stakeholders** because they will differ in interest, contextual knowledge, data literacy, and motives.

[1]HD Lasswell. 1948. The structure and function of communication in society. In *The communication of ideas* (ed. Bryson L), 37-51.

Statistical communication

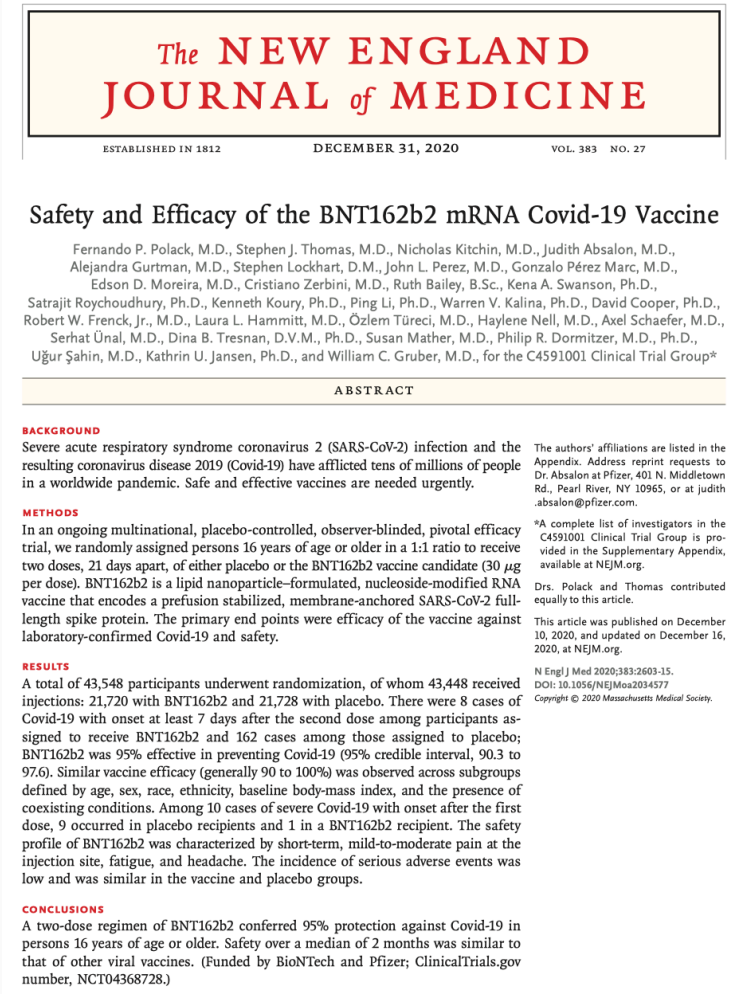
Statistical communication

What we communicate

- The quantity of interest
- The selection/generation of data
- The empirical setup
- The model mechanics and results (estimates/predictions/uncertainty)

Common challenges

- There's epistemological and statistical uncertainty.
- Effect sizes have implications that are often not easy to grasp.
- Conclusions about data science output crucially hinge on the validity of design aspects, which are extremely difficult to communicate.



The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 DECEMBER 31, 2020 VOL. 383 NO. 27

Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine

Fernando P. Polack, M.D., Stephen J. Thomas, M.D., Nicholas Kitchin, M.D., Judith Absalon, M.D., Alejandra Gurtman, M.D., Stephen Lockhart, D.M., John L. Perez, M.D., Gonzalo Pérez Marc, M.D., Edson D. Moreira, M.D., Cristiano Zerbini, M.D., Ruth Bailey, B.Sc., Kena A. Swanson, Ph.D., Satrajit Roychoudhury, Ph.D., Kenneth Koury, Ph.D., Ping Li, Ph.D., Warren V. Kalina, Ph.D., David Cooper, Ph.D., Robert W. Frenck, Jr., M.D., Laura L. Hammitt, M.D., Özlem Türeci, M.D., Haylene Nell, M.D., Axel Schaefer, M.D., Serhat Ünal, M.D., Dina B. Tresnan, D.V.M., Ph.D., Susan Mather, M.D., Philip R. Dormitzer, M.D., Ph.D., Uğur Şahin, M.D., Kathrin U. Jansen, Ph.D., and William C. Gruber, M.D., for the C4591001 Clinical Trial Group*

ABSTRACT

BACKGROUND

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and the resulting coronavirus disease 2019 (Covid-19) have afflicted tens of millions of people in a worldwide pandemic. Safe and effective vaccines are needed urgently.

METHODS

In an ongoing multinational, placebo-controlled, observer-blinded, pivotal efficacy trial, we randomly assigned persons 16 years of age or older in a 1:1 ratio to receive two doses, 21 days apart, of either placebo or the BNT162b2 vaccine candidate (30 μg per dose). BNT162b2 is a lipid nanoparticle–formulated, nucleoside-modified RNA vaccine that encodes a prefusion stabilized, membrane-anchored SARS-CoV-2 full-length spike protein. The primary end points were efficacy of the vaccine against laboratory-confirmed Covid-19 and safety.

RESULTS

A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; BNT162b2 was 95% effective in preventing Covid-19 (95% credible interval, 90.3 to 97.6). Similar vaccine efficacy (generally 90 to 100%) was observed across subgroups defined by age, sex, race, ethnicity, baseline body-mass index, and the presence of coexisting conditions. Among 10 cases of severe Covid-19 with onset after the first dose, 9 occurred in placebo recipients and 1 in a BNT162b2 recipient. The safety profile of BNT162b2 was characterized by short-term, mild-to-moderate pain at the injection site, fatigue, and headache. The incidence of serious adverse events was low and was similar in the vaccine and placebo groups.

CONCLUSIONS

A two-dose regimen of BNT162b2 conferred 95% protection against Covid-19 in persons 16 years of age or older. Safety over a median of 2 months was similar to that of other viral vaccines. (Funded by BioNTech and Pfizer; ClinicalTrials.gov number, NCT04368728.)

The authors' affiliations are listed in the Appendix. Address reprint requests to Dr. Absalon at Pfizer, 401 N. Middletown Rd., Pearl River, NY 10965, or at judith.absalon@pfizer.com.

*A complete list of investigators in the C4591001 Clinical Trial Group is provided in the Supplementary Appendix, available at NEJM.org.

Drs. Polack and Thomas contributed equally to this article.

This article was published on December 10, 2020, and updated on December 16, 2020, at NEJM.org.

N Engl J Med 2020;383:2603-15.
DOI: 10.1056/NEJMoa2034577

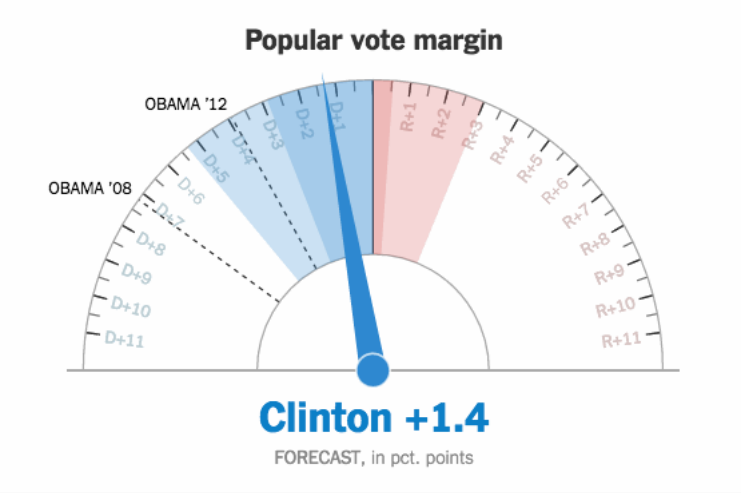
`Credit` Polack et al. 2020, NEJM

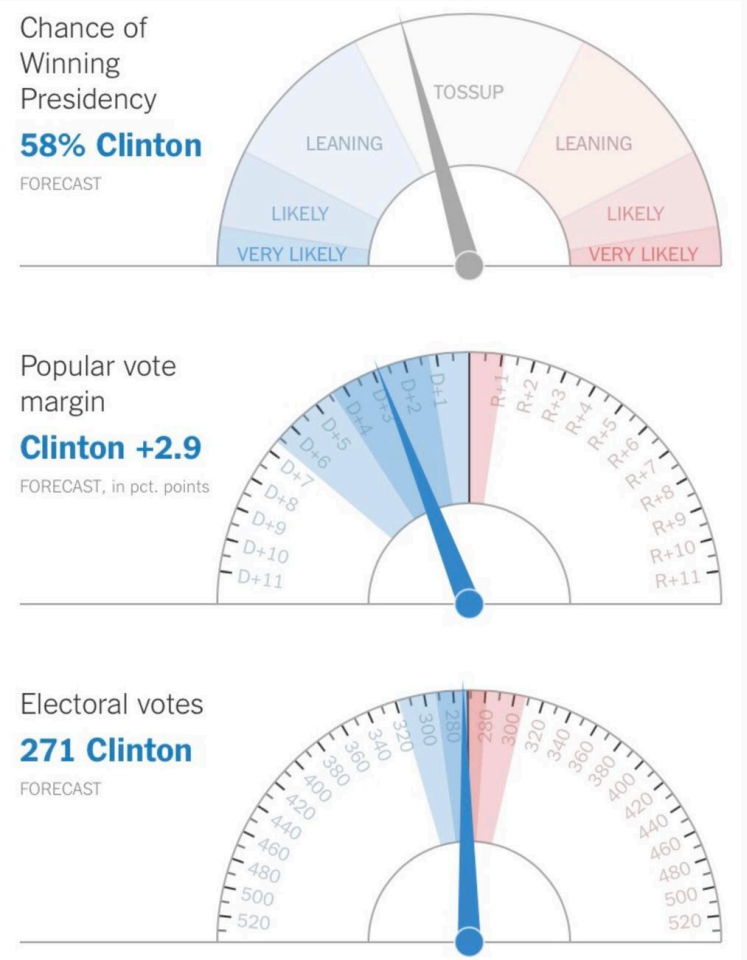
Uncertainty

Hertie School

Question to reflect on

Thinking about the task of publicly forecasting the U.S. presidential election using polling data, what kinds of uncertainty would you have to deal with as a forecaster?

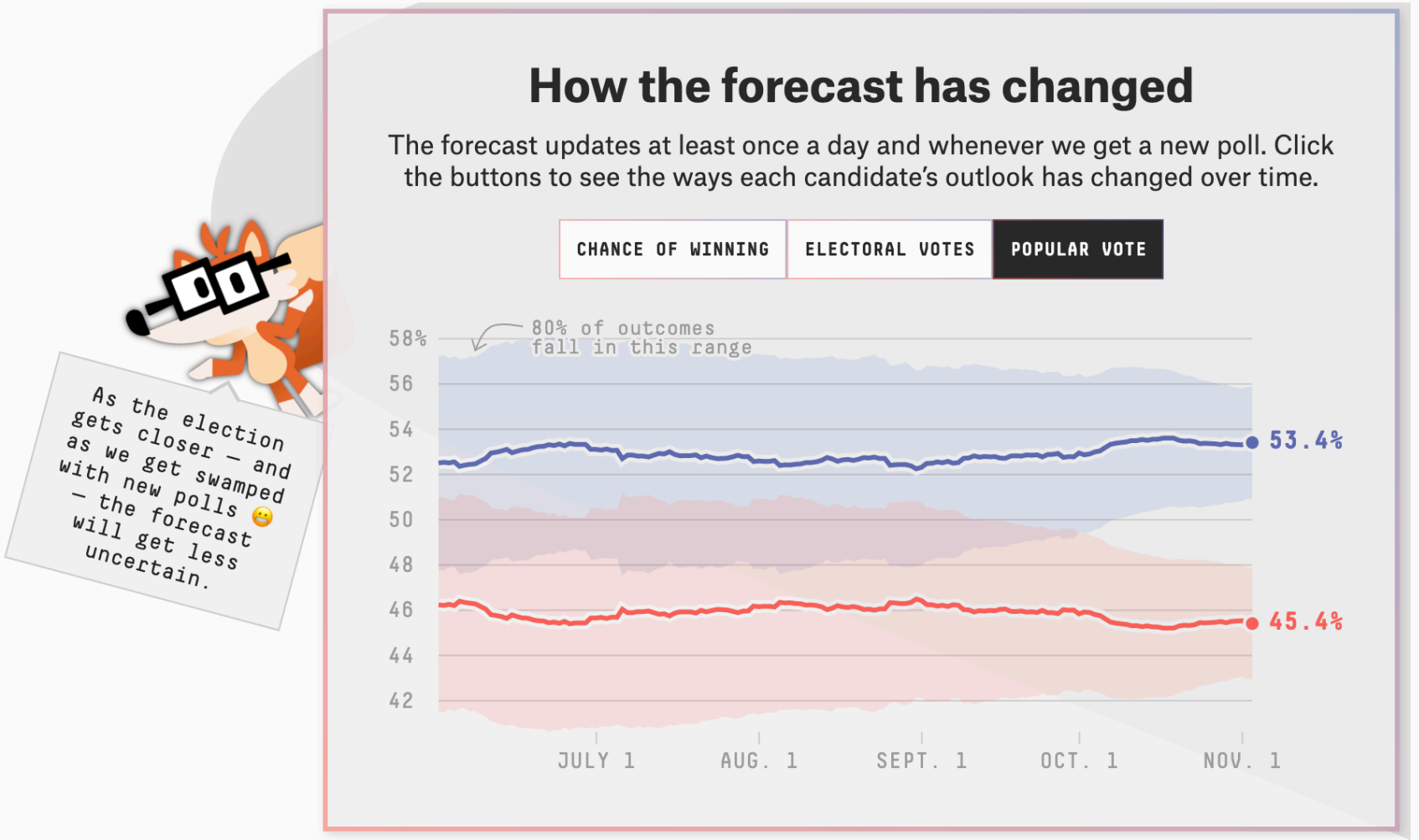




`Credit` NYTimes.com at 9:20 p.m. Nov. 8, 2016

Example: FiveThirtyEight 2020 election forecast

Hertie School



Source FiveThirtyEight

Example: What is behind the visuals?

The components of our uncertainty index are as follows:

1. The number of undecided voters in national polls. More undecided voters means more uncertainty.
2. The number of undecided plus third-party voters in national polls. More third-party voters means more uncertainty.
3. Polarization, as measured elsewhere in the model, is based on how far apart the parties are in roll call votes cast in the U.S. House. More polarization means less uncertainty since there are fewer swing voters.
4. The volatility of the national polling average. Volatility tends to predict itself, so a stable polling average tends to remain stable.
5. The overall volume of national polling. More polling means less uncertainty.
6. The magnitude of the difference between the polling-based national snapshot and the fundamentals forecast. A wider gap means more uncertainty.
7. The standard deviation of the component variables used in the FiveThirtyEight economic index. More economic volatility means more overall uncertainty in the forecast.
8. The volume of major news, as measured by the number of full-width New York Times headlines in the past 500 days, with more recent days weighted more heavily. More news means more uncertainty.

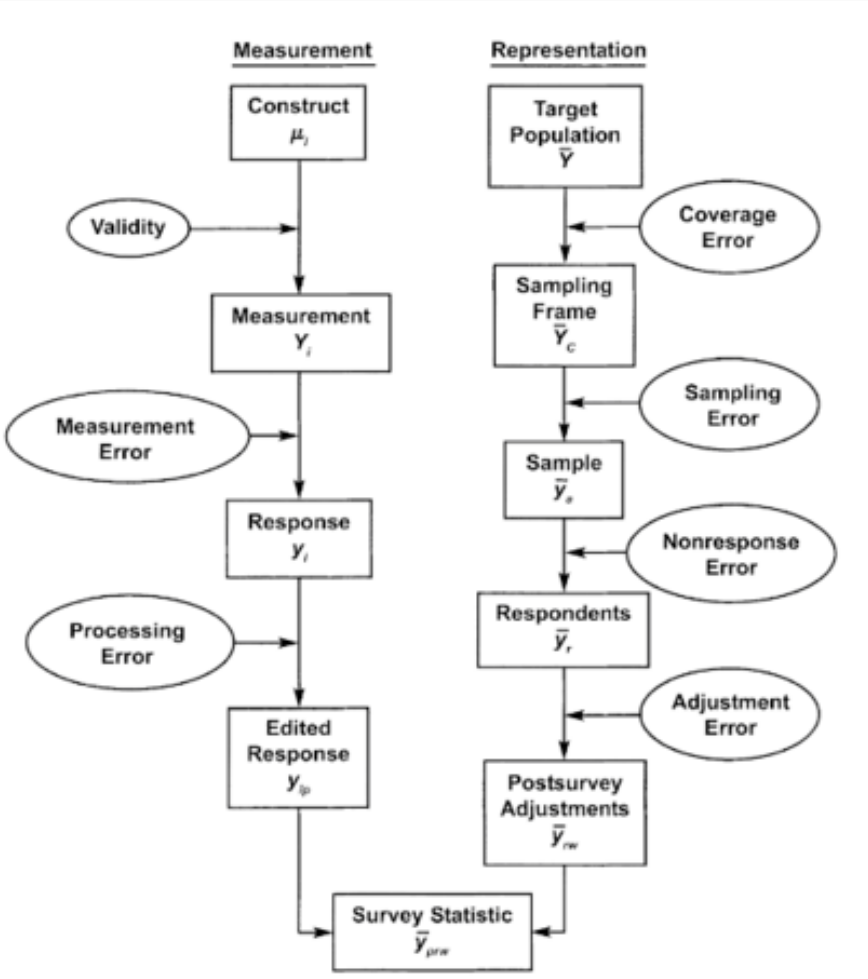
`Source` FiveThirtyEight

Uncertainty

What we are uncertain about

- **Measurement** → uncertainty in single variables
- **Model specification** → uncertainty across multiple variables and how they connect
- **Parameter estimates** → uncertainty about bias and precision
- **Model outcomes** → uncertainty about (out-of-sample) fit
- **Generalizability** to other samples, the future

Depending on the empirical setup, various specific **sources of error** might enter (e.g., survey data, digital trace data).



Credit Robert Groves, Total Survey Error

Communicating uncertainty

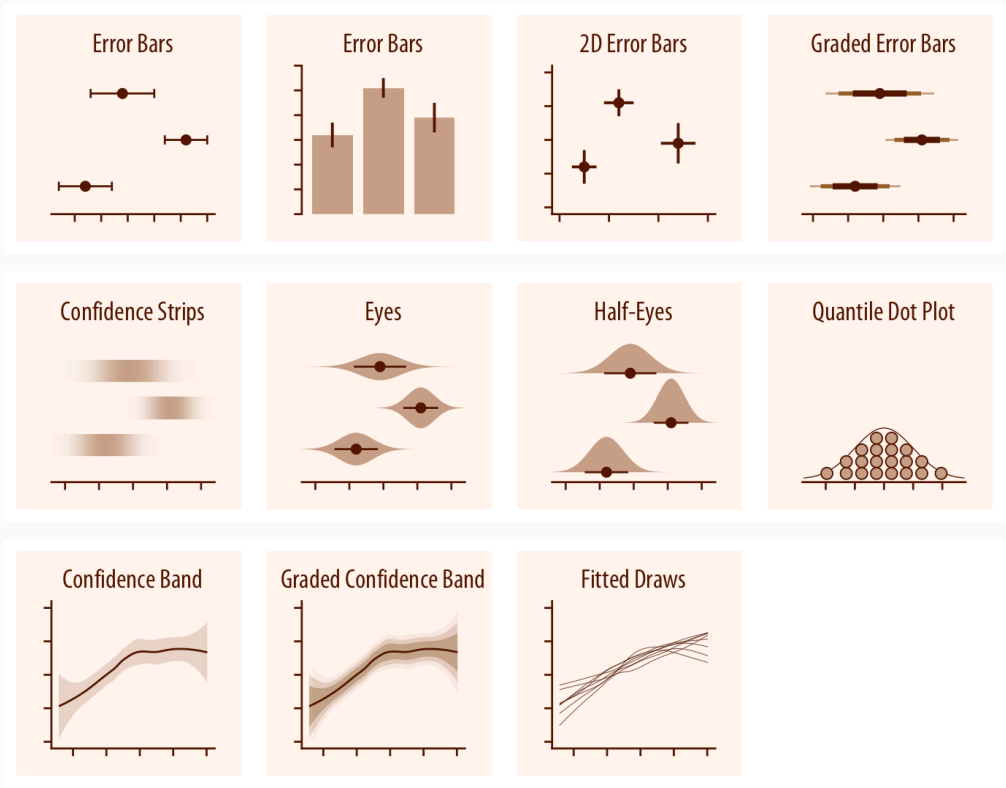
The difficulty of communicating uncertainty

- The concept is **complex**. Not all people think in probabilistic terms.
- **Many humans are bad at understanding (conditional and unconditional) probabilities**.
- Adding **information about uncertainty might distract**, confuse, and undermine trust.

`Credit` Richard McElreath

Communicating uncertainty (cont.)

Visualizing uncertainty

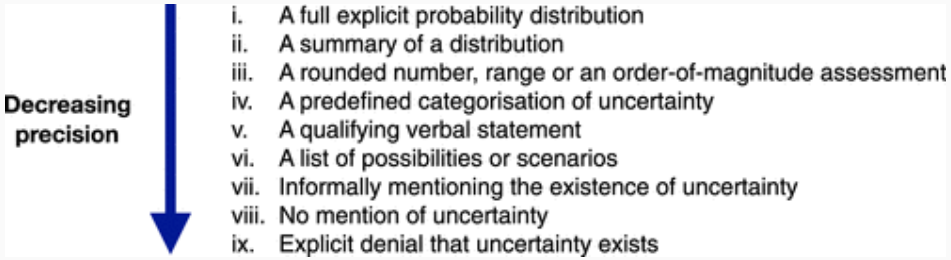


`Source` Claus Wilke

Uncertainty by numbers

```
> broom::tidy(model_out, conf.int = TRUE, conf.level = 0.95)
# A tibble: 4 × 7
  term        estimate std.error statistic   p.value conf.low conf.high
  <chr>          <dbl>     <dbl>     <dbl>     <dbl>    <dbl>     <dbl>
1 (Intercept) 13.4      0.175        76.7  0          13.1      13.8
2 distance    -0.00405  0.000110    -36.9  5.53e-297 -0.00426  -0.00383
3 originJFK   -2.70     0.189       -14.3  1.46e- 46 -3.07     -2.33
4 originLGA   -4.46     0.194       -23.0  3.04e-117 -4.84     -4.08
```

Strategies by precision



`Credit` van der Bles et al. 2019

Communicating probabilities in text form

Hertie School

JCOM

Variability in the interpretation of probability phrases used in Dutch news articles — a risk for miscommunication

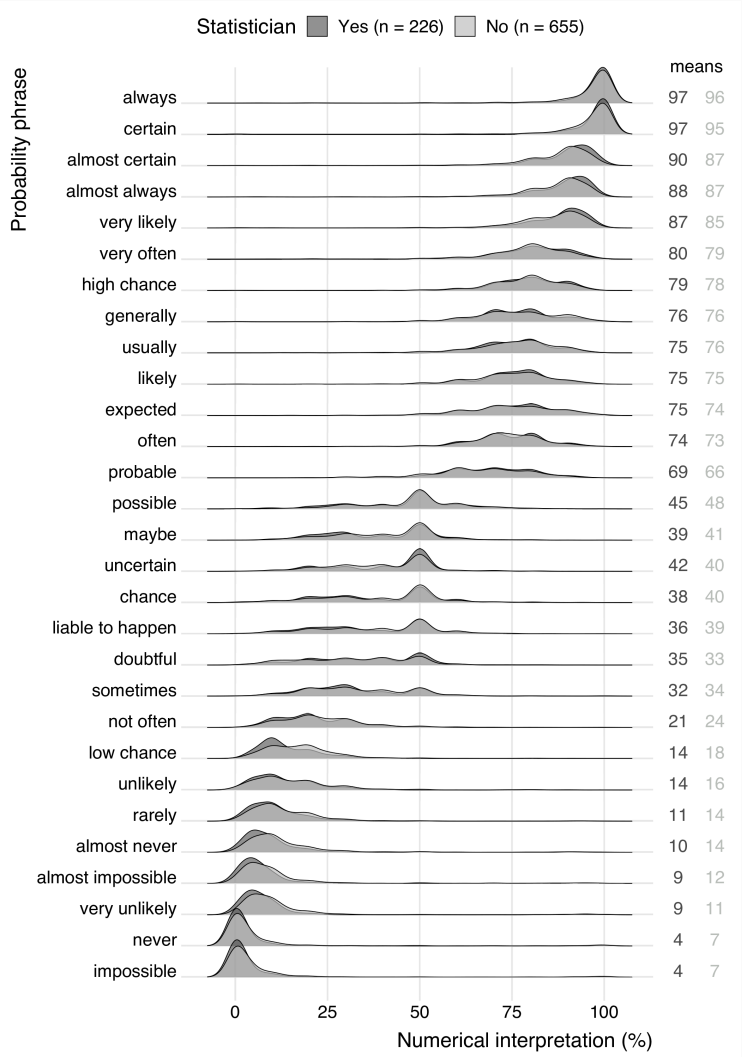
Sanne Willems, Casper Albers and Ionica Smeets

Abstract

Verbal probability phrases are often used in science communication to express estimated risks in words instead of numbers. In this study we look at how laypeople and statisticians interpret Dutch probability phrases that are regularly used in news articles. We found that there is a large variability in interpretations, even if the phrases are given in a neutral context. Also, statisticians do not agree on the interpretation of the phrases. We conclude that science communicators should be careful in using verbal probability expressions.

Keywords Risk communication; Science and media; Science writing

`Source` Willems et al. 2020



Written communication in data science workflows

When to communicate in writing

- For communicating to **the public and decision makers**, who want to focus on the conclusions, not the code behind the analysis.
- For collaborating with **other data scientists**, who are interested in both your conclusions and how you reached them (i.e. the code).

Written communication

When to communicate in writing

- For communicating to **the public and decision makers**, who want to focus on the conclusions, not the code behind the analysis.
- For collaborating with **other data scientists**, who are interested in both your conclusions and how you reached them (i.e. the code).

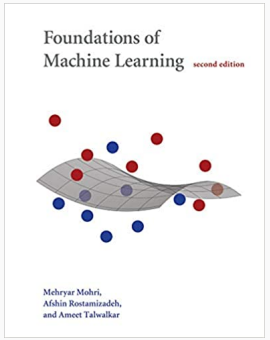
Authoring as part of the workflow

- Many different formats, including reports, briefs, blog posts, books, presentations, ...
- *Form follows function*: the write-up tool should talk to the analytic toolset.

The continuum of written data science communication

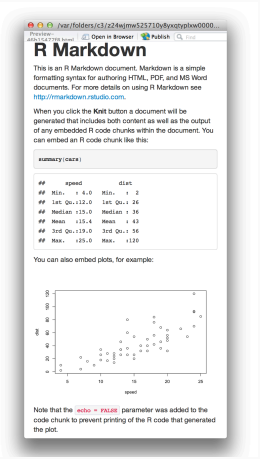
Hertie School

Book



Conference proceeding, journal article

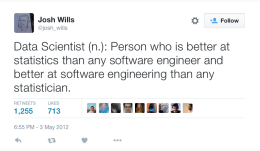
Technical report



Executive summary

Dashboard

Tweet



Programmatic solutions for communicating

Open-source software can help

- R Markdown (and the `rmarkdown` package) helps you create dynamic analysis documents that combine code, rendered output (such as figures), and prose.
- You can use it to
  - Do data science interactively with notebooks.
  - Modify the layout of your report.
  - Communicate your results with others.
- You take care of content, R Markdown of format.

More resources

- The official website
- The R Markdown Cookbook
- R Markdown - The Definitive Guide

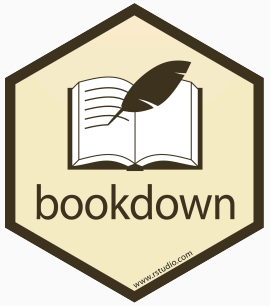
The possibilities for communicating in data sicence workflows are immense

- R Markdown can do much more than reports. You can use it to author
  - Reports (in PDF, HTML, Word, etc.)
  - Interactive documents
  - Dashboards
  - Slideshows
  - Books
  - Websites
- It stands on the shoulders of Pandoc, a program that converts markup files into virtually any other format.

RMarkdown formats

bookdown

- A package that facilitates writing books and long-form articles/reports with R Markdown.
- See here for an overview of books written with `bookdown`.



blogdown

- A package that lets you create websites (not only blogs!) using R Markdown.
- It integrates Hugo (or other site generators).

pagedown

- A package that lets you paginate the HTML output of R Markdown with CSS for print (PDF).
- Lots of different templates available.

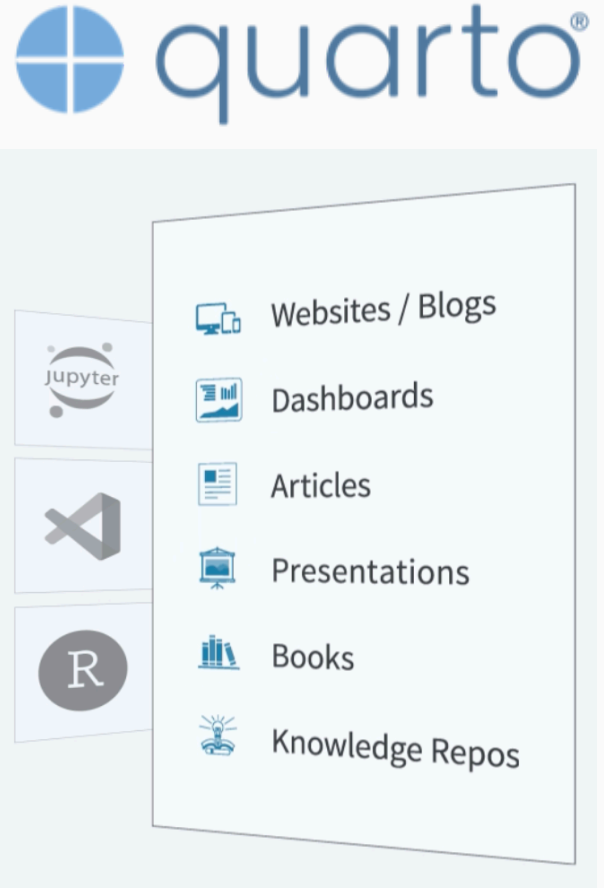
xaringan

- A package that lets you crate slideshows with remark.js through R Markdown.
- These slides have been created using this package.

The next generation of technical publishing: Quarto

What's Quarto?

- Quarto is software developed at Posit, the company behind RStudio.
- It's the "next generation of R Markdown" and also built on Pandoc. If you know R Markdown well, you already know Quarto well.
- It facilitates embedding code and output from R, Python, Julia, and other languages.
- It combines the functionality of R Markdown and all the other tools (bookdown, xaringan, etc.) into one single consistent system.
- Quarto is still fairly new and under active development. More and more extensions are coming out that increase the flexibility of the suite.
- Check out the comprehensive guide to learn more.



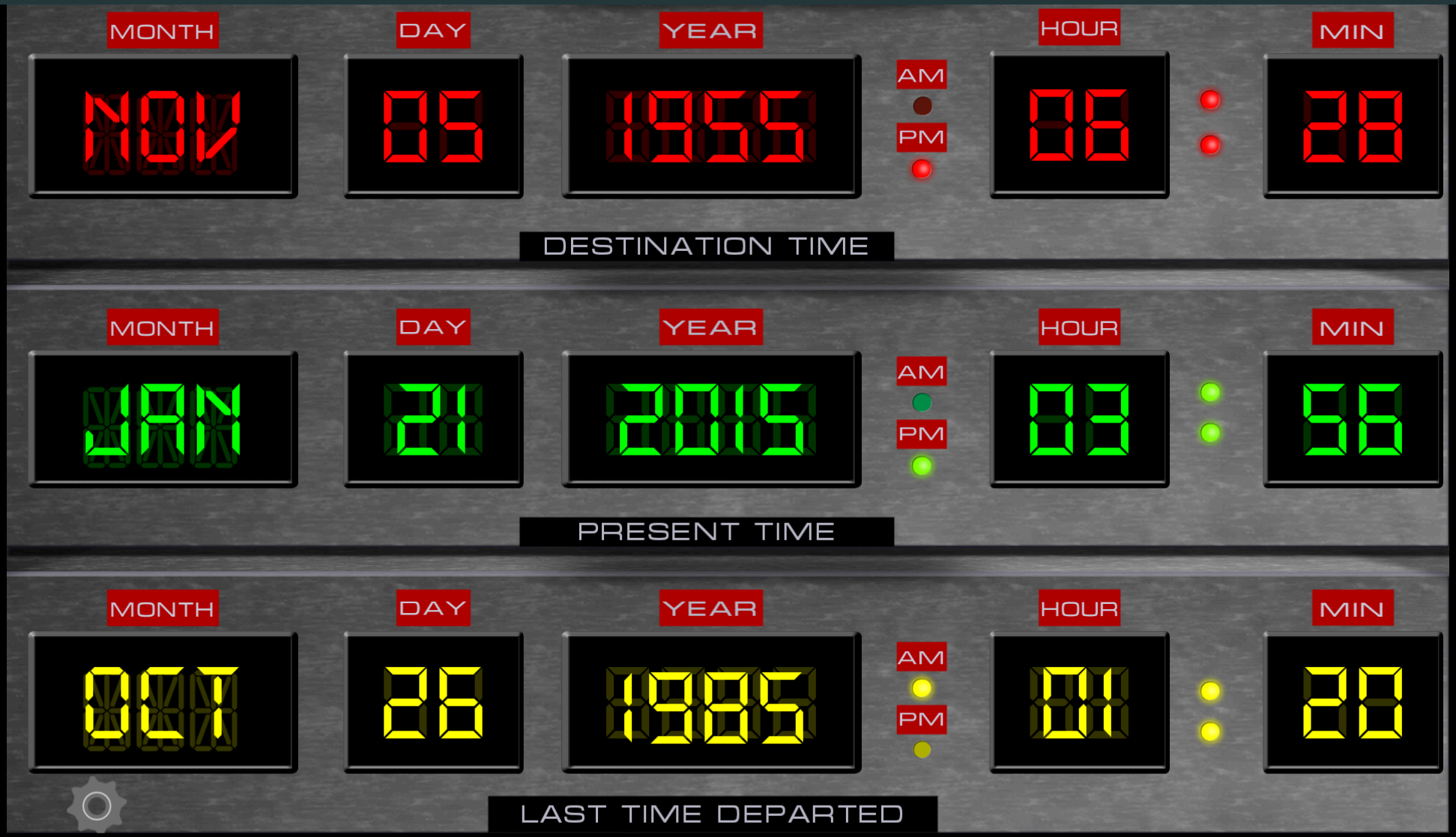
Quarto and Jupyter

- Quarto's support for both Knitr and Jupyter means that you can use it to create documentation projects that contain content from both systems.
- Something to keep in mind next semester and going forward! :)

Interactive communication with dashboards

What are dashboards?

Hertie School

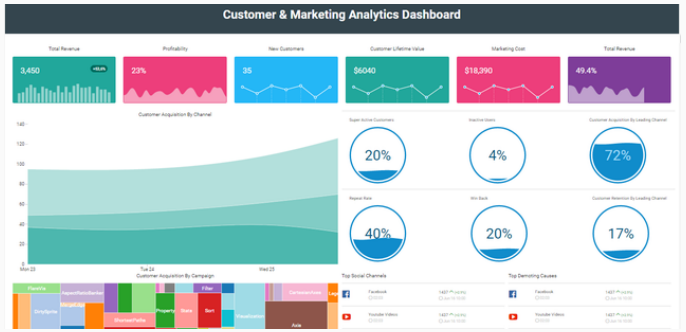


What are dashboards?

Yeah, what are they really?

- A (business or data) dashboard is a GUI that provides high-level overviews of performance indicators or other quantities of interest.

`Credit` Tim Green



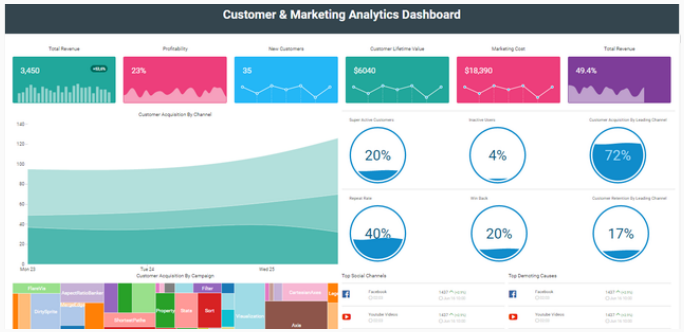
`Credit` HelicalInsight OpenSourceBI

What are dashboards?

Yeah, what are they really?

- A (business or data) dashboard is a GUI that provides high-level overviews of performance indicators or other quantities of interest.

`Credit` Tim Green



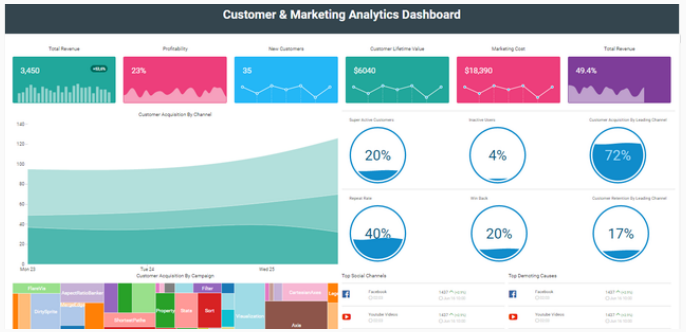
`Credit` HelicalInsight OpenSourceBI

What are dashboards?

Yeah, what are they really?

- A (business or data) dashboard is a GUI that provides high-level overviews of performance indicators or other quantities of interest.
- It's a **monitoring** (and not so much analysis) **tool**.

`Credit` Tim Green



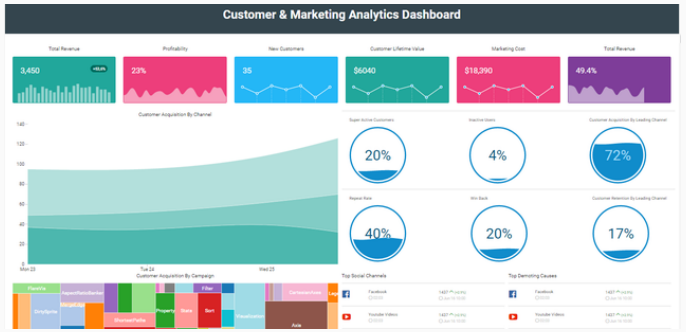
`Credit` HelicalInsight OpenSourceBI

What are dashboards?

Yeah, what are they really?

- A (business or data) dashboard is a GUI that provides high-level overviews of performance indicators or other quantities of interest.
- It's a **monitoring** (and not so much analysis) **tool**.
- Think of dashboards as a **mash-up of data visualization and report**.

`Credit` Tim Green



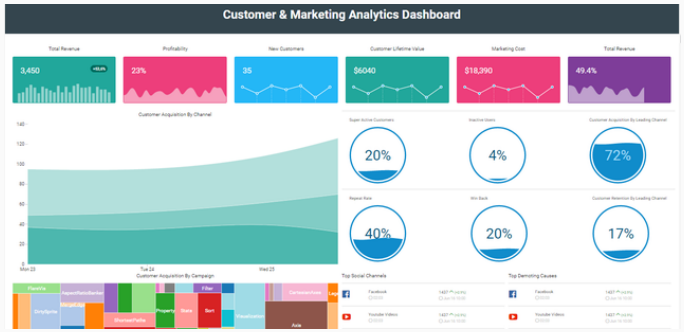
`Credit` HelicalInsight OpenSourceBI

What are dashboards?

Yeah, what are they really?

- A (business or data) dashboard is a GUI that provides high-level overviews of performance indicators or other quantities of interest.
- It's a **monitoring** (and not so much analysis) **tool**.
- Think of dashboards as a **mash-up of data visualization and report**.
- Dashboards are increasingly popular in businesses and organizations to **synthesize data points** from operative units (for strategic and analytical purposes).

`Credit` Tim Green



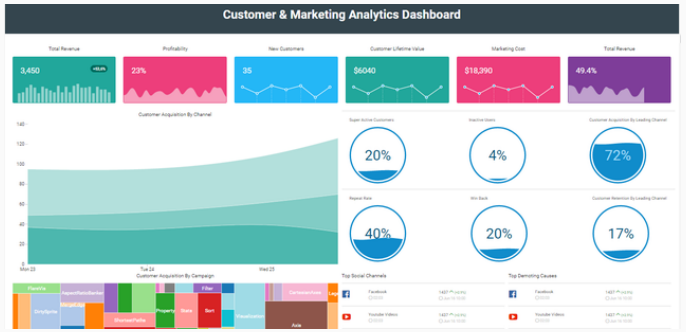
`Credit` HelicalInsight OpenSourceBI

What are dashboards?

Yeah, what are they really?

- A (business or data) dashboard is a GUI that provides high-level overviews of performance indicators or other quantities of interest.
- It's a **monitoring** (and not so much analysis) **tool**.
- Think of dashboards as a **mash-up of data visualization and report**.
- Dashboards are increasingly popular in businesses and organizations to **synthesize data points** from operative units (for strategic and analytical purposes).
- Data journalism has started to embrace dashboards in the context of elections, the COVID-19 pandemic, and sports.

`Credit` Tim Green



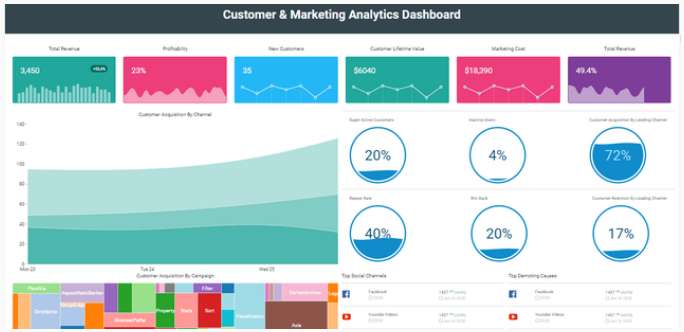
`Credit` HelicalInsight OpenSourceBI

What are dashboards?

Yeah, what are they really?

- A (business or data) dashboard is a GUI that provides high-level overviews of performance indicators or other quantities of interest.
- It's a **monitoring** (and not so much analysis) **tool**.
- Think of dashboards as a **mash-up of data visualization and report**.
- Dashboards are increasingly popular in businesses and organizations to **synthesize data points** from operative units (for strategic and analytical purposes).
- Data journalism has started to embrace dashboards in the context of elections, the COVID-19 pandemic, and sports.
- **Common features** are:
  - Accessibility via web browser
  - Featuring of interactives
  - Heavy focus on comparative visualization
  - Provision of trends on key performance indicators (KPIs)

`Credit` Tim Green

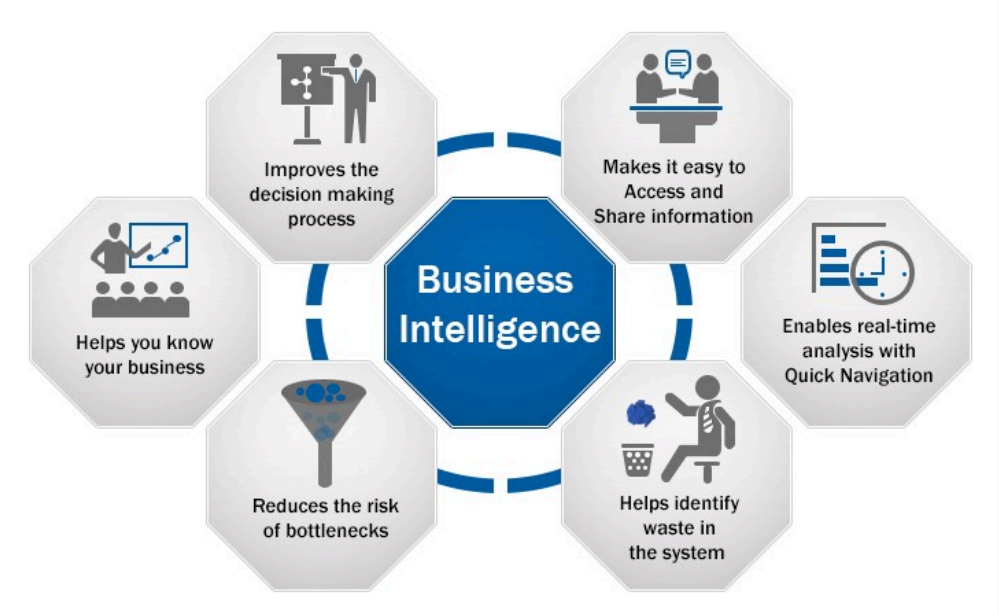


`Credit` HelicalInsight OpenSourceBI

Why are dashboards?

Why are they a thing?

- There is increasing **abundance of data** (often process-generated) that cannot speak for itself.

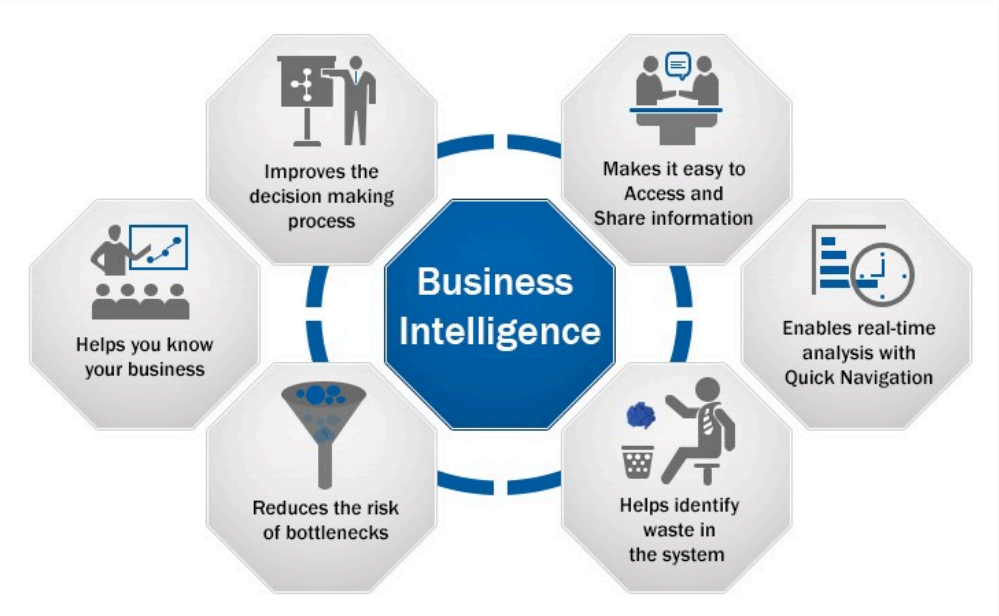


`Credit` towardsdatascience.com

Why are dashboards?

Why are they a thing?

- There is increasing **abundance of data** (often process-generated) that cannot speak for itself.
- If used wisely, these data can provide an **important part of business intelligence** and a basis for high-level evidence-based decision-making.

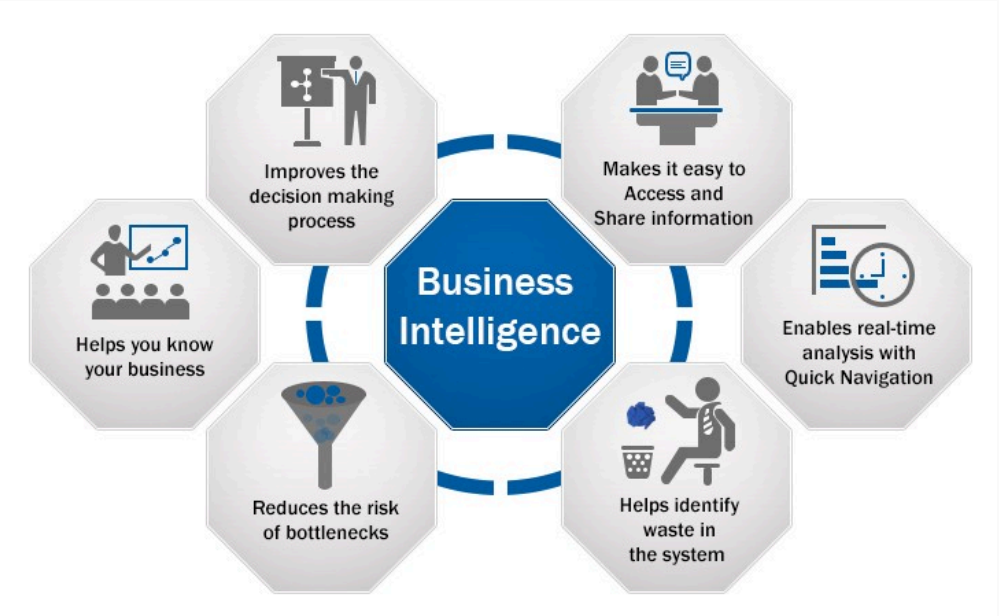


`Credit` towardsdatascience.com

Why are dashboards?

Why are they a thing?

- There is increasing **abundance of data** (often process-generated) that cannot speak for itself.
- If used wisely, these data can provide an **important part of business intelligence** and a basis for high-level evidence-based decision-making.
- Provide continuous quantification of indicators of interest (→ **monitoring**).

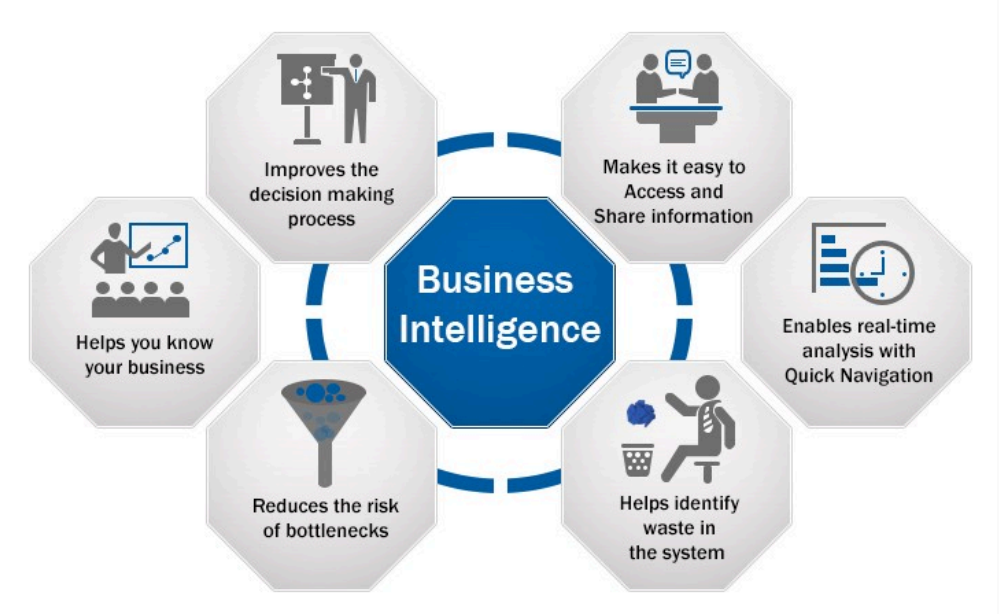


`Credit` towardsdatascience.com

Why are dashboards?

Why are they a thing?

- There is increasing **abundance of data** (often process-generated) that cannot speak for itself.
- If used wisely, these data can provide an **important part of business intelligence** and a basis for high-level evidence-based decision-making.
- Provide continuous quantification of indicators of interest (→ **monitoring**).
- **Reduce information differential** between analysts and stakeholders.

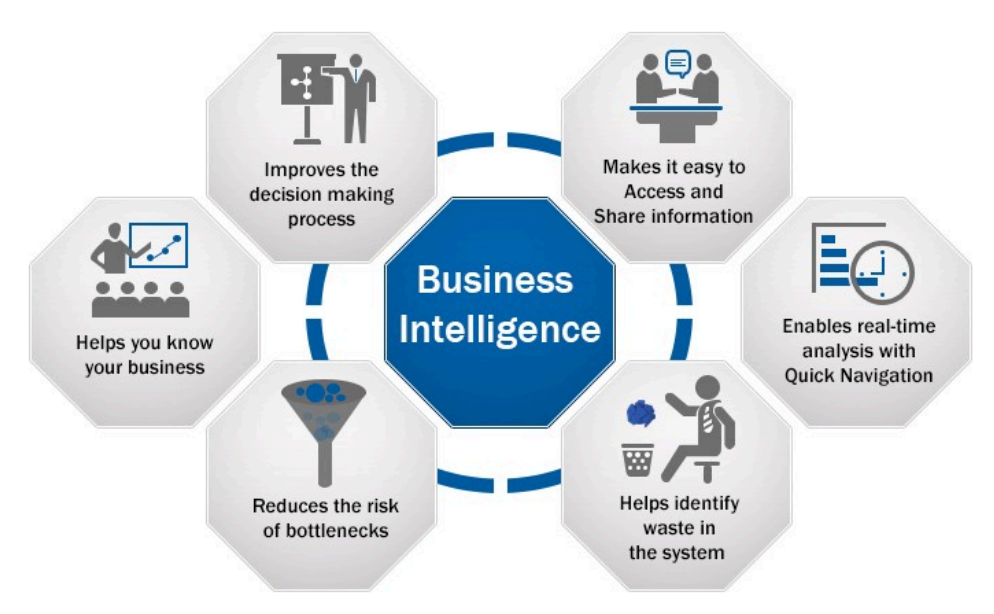


`Credit` towardsdatascience.com

Why are dashboards?

Why are they a thing?

- There is increasing **abundance of data** (often process-generated) that cannot speak for itself.
- If used wisely, these data can provide an **important part of business intelligence** and a basis for high-level evidence-based decision-making.
- Provide continuous quantification of indicators of interest (→ **monitoring**).
- **Reduce information differential** between analysts and stakeholders.
- Also, **measuring the health of organizations** can help stay in control (if only as a performative act) and satisfy managers' need for micromanaging.



`Credit` towardsdatascience.com

Dashboards in the wild

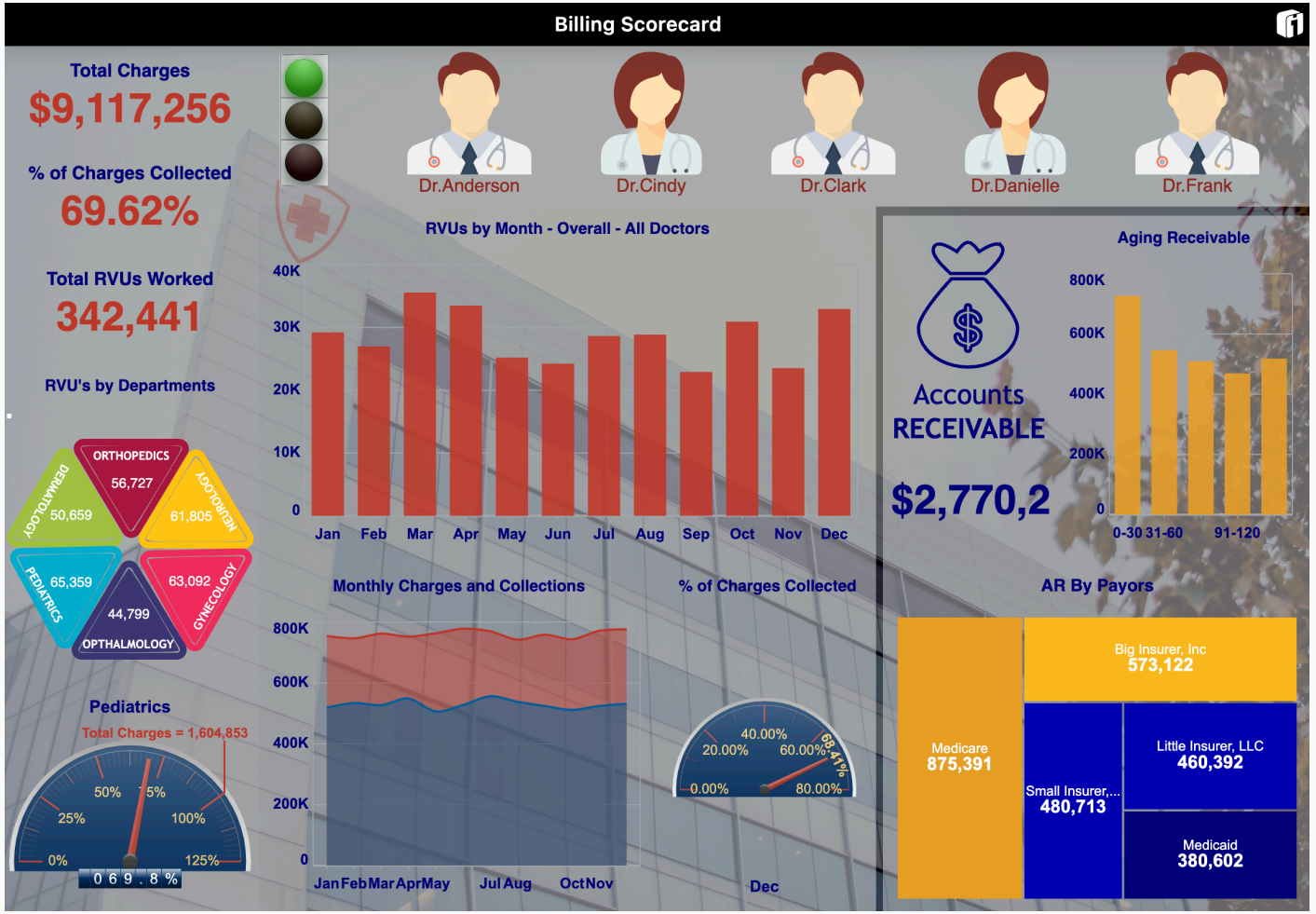
`Credit` geckoboard.com

Dashboards in the wild

Hertie School

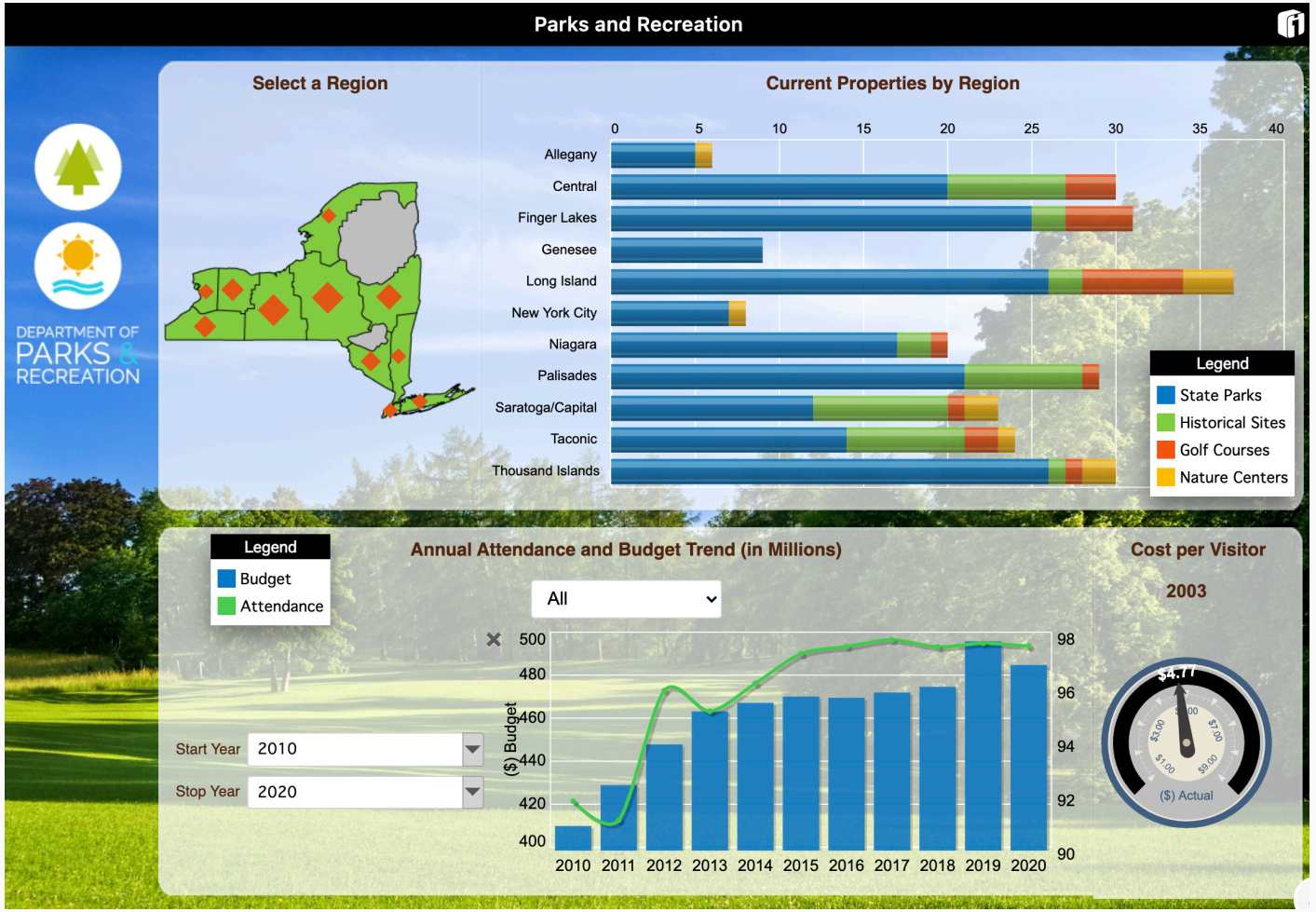
`Credit` geckoboard.com

Dashboards in the wild



Credit idashboards.com

Dashboards in the wild

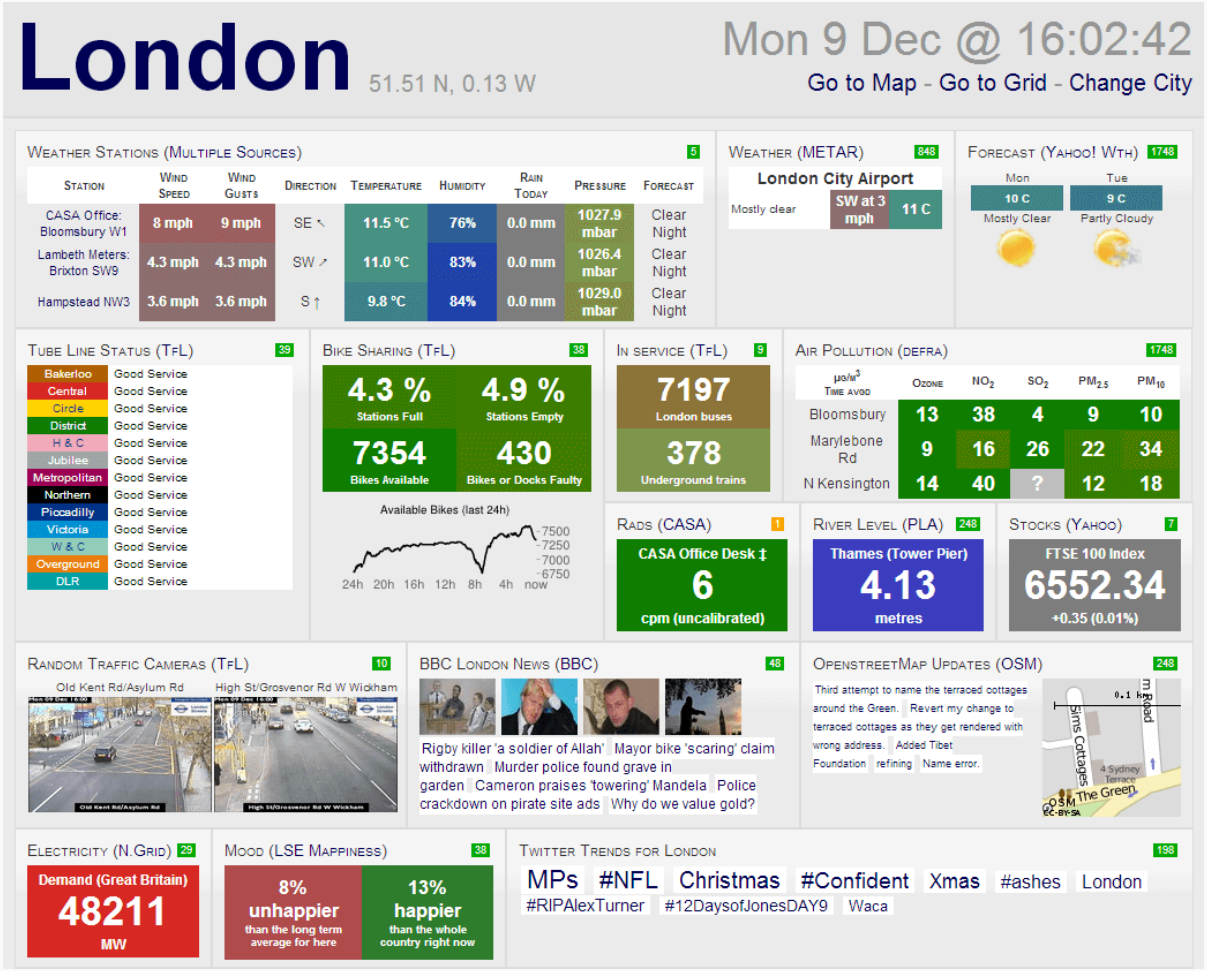


Credit idashboards.com

Dashboards in the wild

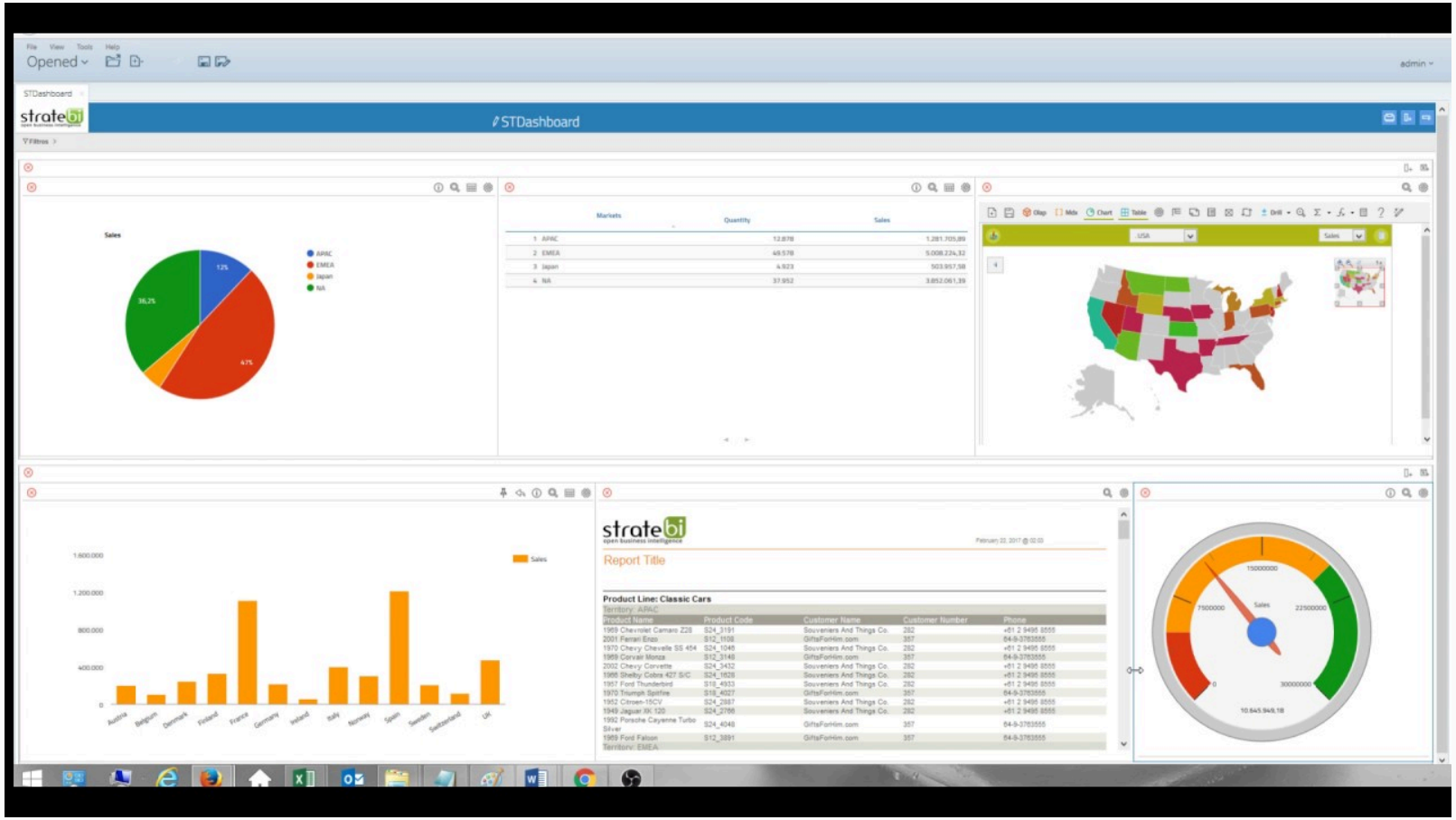
Credit Stephen Few

Dashboards in the wild



Credit idashboards.com

Dashboards in the wild



`Credit` carmel.es

The problem(s) with dashboards

Design challenges

- **They say too little.** Loss of information is fatal for good decision-making when aggregating results into few KPIs.
- **They say too much** (irrelevant things).
- Dashboards often fail not in technology but in communication (rooted in poor design).
- "Dashboards are not for show. No amount of cuteness and technical wizardry can substitute for clear communication." Stephen Few, Perceptual Edge
- Dashboards are a subgenre of data viz, so **all rules of good/bad viz apply**.
- So, there is hope since we do know a bit about how to design good visuals. (See here for a nice case study on improving the design of a dashboard.)

Analytic challenges

- Dashboards cater to the desire to be able to make good decisions on the basis of few selected metrics.
- This logic reflects a **gross simplification of reality**.
- All challenges that pop up in careful analytic work - issues of selection, measurement, causality, predictiveness - are still valid but will be obscured when aggregating data.
- Simple metrics can still be useful, but often **you need contextual knowledge** (which is difficult to communicate in dashboards).
- Another consequence of "dashboarding" business intelligence can be that by making decisions a function of metrics, they stop working well because they will be gamed.

Thoughtful dashboard design and usage

Checklist before you start[1]

1. Are you tackling a monitoring task that needs your data/metrics to be updated frequently?
2. Who will use the dashboard and to what end? What questions will they use it to answer? What actions will they take in response to these answers?
3. What specific information should be displayed, and is it meaningful without much context?
4. What could lead to the metrics being wrong/misleading?

Design advice

- Minimize distractions.

[1]Source: Stephen Few/Perceptual Edge

Thoughtful dashboard design and usage

Checklist before you start[1]

1. Are you tackling a monitoring task that needs your data/metrics to be updated frequently?
2. Who will use the dashboard and to what end? What questions will they use it to answer? What actions will they take in response to these answers?
3. What specific information should be displayed, and is it meaningful without much context?
4. What could lead to the metrics being wrong/misleading?

Design advice

- Minimize distractions.
- Focus on meaningful quantities of interest, not the ones that look cool.

[1]Source: Stephen Few/Perceptual Edge

Thoughtful dashboard design and usage

Checklist before you start[1]

1. Are you tackling a monitoring task that needs your data/metrics to be updated frequently?
2. Who will use the dashboard and to what end? What questions will they use it to answer? What actions will they take in response to these answers?
3. What specific information should be displayed, and is it meaningful without much context?
4. What could lead to the metrics being wrong/misleading?

Design advice

- Minimize distractions.
- Focus on meaningful quantities of interest, not the ones that look cool.
- Don't overload with information.

[1]Source: Stephen Few/Perceptual Edge

Thoughtful dashboard design and usage

Checklist before you start[1]

1. Are you tackling a monitoring task that needs your data/metrics to be updated frequently?
2. Who will use the dashboard and to what end? What questions will they use it to answer? What actions will they take in response to these answers?
3. What specific information should be displayed, and is it meaningful without much context?
4. What could lead to the metrics being wrong/misleading?

Design advice

- Minimize distractions.
- Focus on meaningful quantities of interest, not the ones that look cool.
- Don't overload with information.
- Apply all rules of good data viz.

[1]Source: Stephen Few/Perceptual Edge

Thoughtful dashboard design and usage

Checklist before you start[1]

1. Are you tackling a monitoring task that needs your data/metrics to be updated frequently?
2. Who will use the dashboard and to what end? What questions will they use it to answer? What actions will they take in response to these answers?
3. What specific information should be displayed, and is it meaningful without much context?
4. What could lead to the metrics being wrong/misleading?

Design advice

- Minimize distractions.
- Focus on meaningful quantities of interest, not the ones that look cool.
- Don't overload with information.
- Apply all rules of good data viz.
- Use interactives with care (e.g., to make optional content conditionally visible)

[1]Source: Stephen Few/Perceptual Edge

Thoughtful dashboard design and usage

Checklist before you start[1]

1. Are you tackling a monitoring task that needs your data/metrics to be updated frequently?
2. Who will use the dashboard and to what end? What questions will they use it to answer? What actions will they take in response to these answers?
3. What specific information should be displayed, and is it meaningful without much context?
4. What could lead to the metrics being wrong/misleading?

Design advice

- Minimize distractions.
- Focus on meaningful quantities of interest, not the ones that look cool.
- Don't overload with information.
- Apply all rules of good data viz.
- Use interactives with care (e.g., to make optional content conditionally visible)
- Try not to exceed the boundaries of a single screen.

[1]Source: Stephen Few/Perceptual Edge

Thoughtful dashboard design and usage

Checklist before you start[1]

1. Are you tackling a monitoring task that needs your data/metrics to be updated frequently?
2. Who will use the dashboard and to what end? What questions will they use it to answer? What actions will they take in response to these answers?
3. What specific information should be displayed, and is it meaningful without much context?
4. What could lead to the metrics being wrong/misleading?

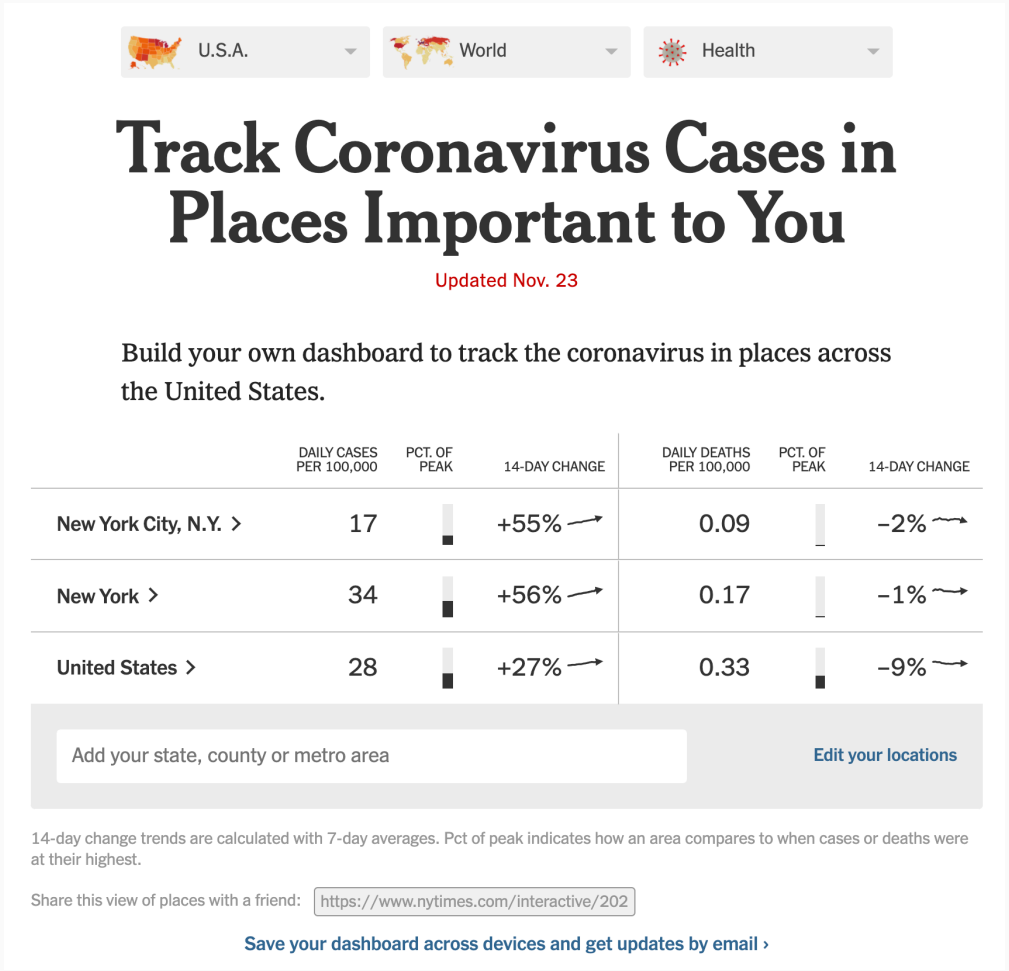
[1]Source: Stephen Few/Perceptual Edge

Design advice

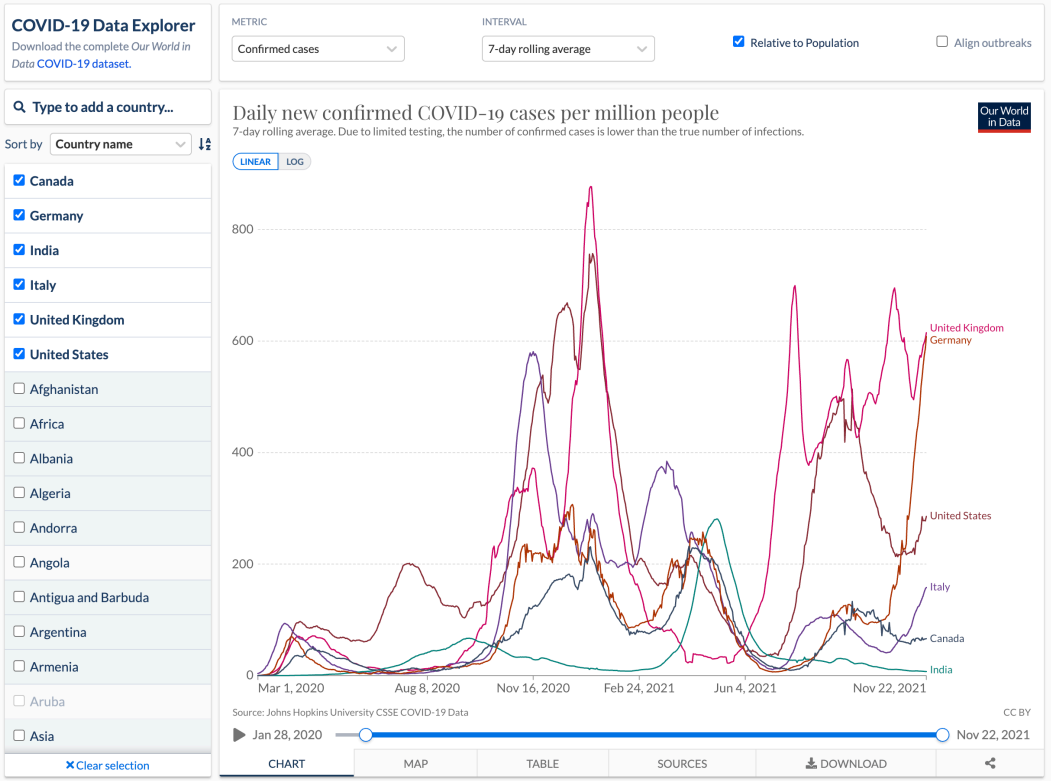
- Minimize distractions.
- Focus on meaningful quantities of interest, not the ones that look cool.
- Don't overload with information.
- Apply all rules of good data viz.
- Use interactives with care (e.g., to make optional content conditionally visible)
- Try not to exceed the boundaries of a single screen.
- Ensure desktop/mobile screen responsiveness.

Dashboards in the wild: COVID-19 edition

Hertie School



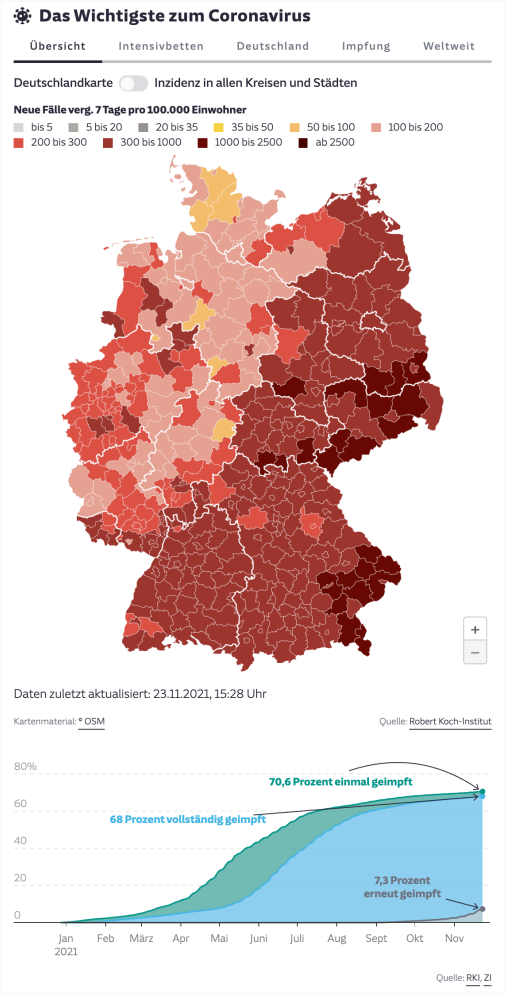
Credit NY Times



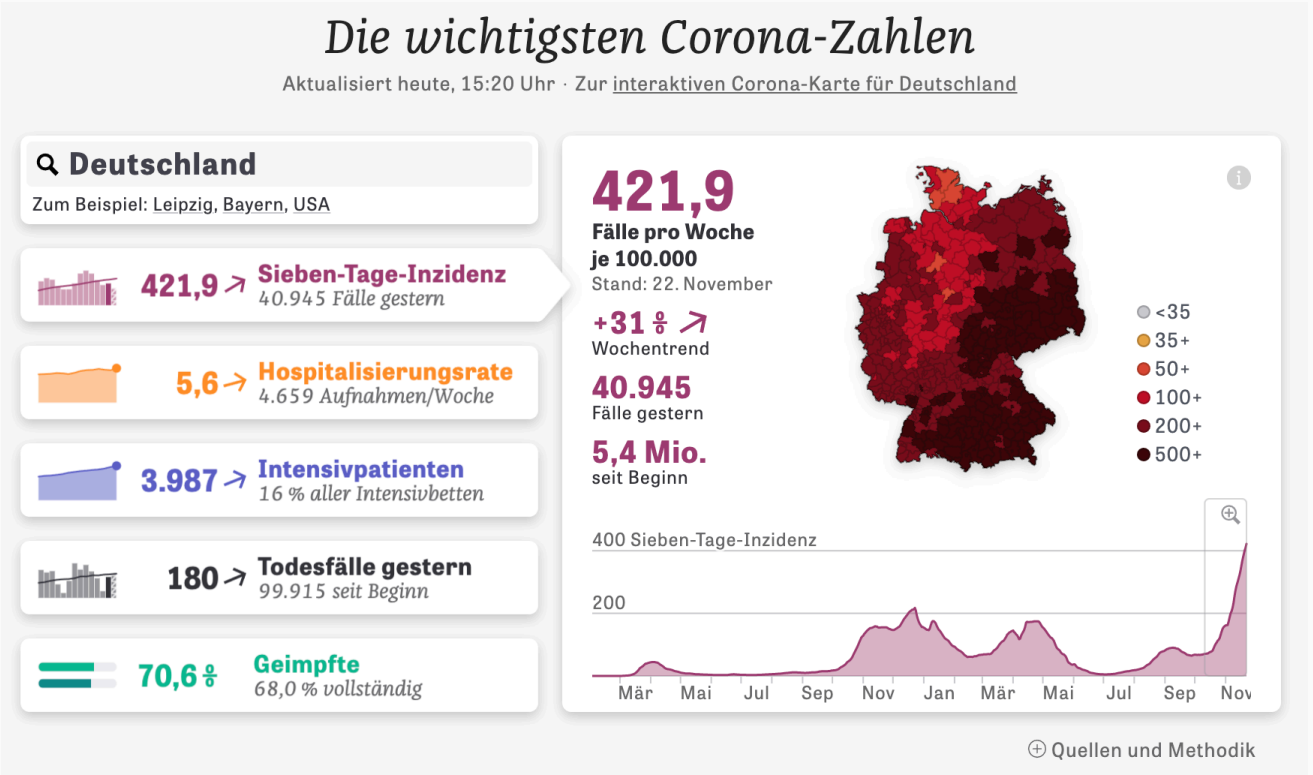
Credit Our World in Data

Dashboards in the wild: COVID-19 edition

Hertie School



Credit SZ Online



Credit ZEIT Online

Dashboards: another example

Hertie School

Schools | Regions

COLOR SCHOOLS

Average score (/300)

FILTER SCHOOLS

Context (by council)

Urban Rural

Regions (26)

ALL REGIONS

Councils (184)

ALL COUNCILS

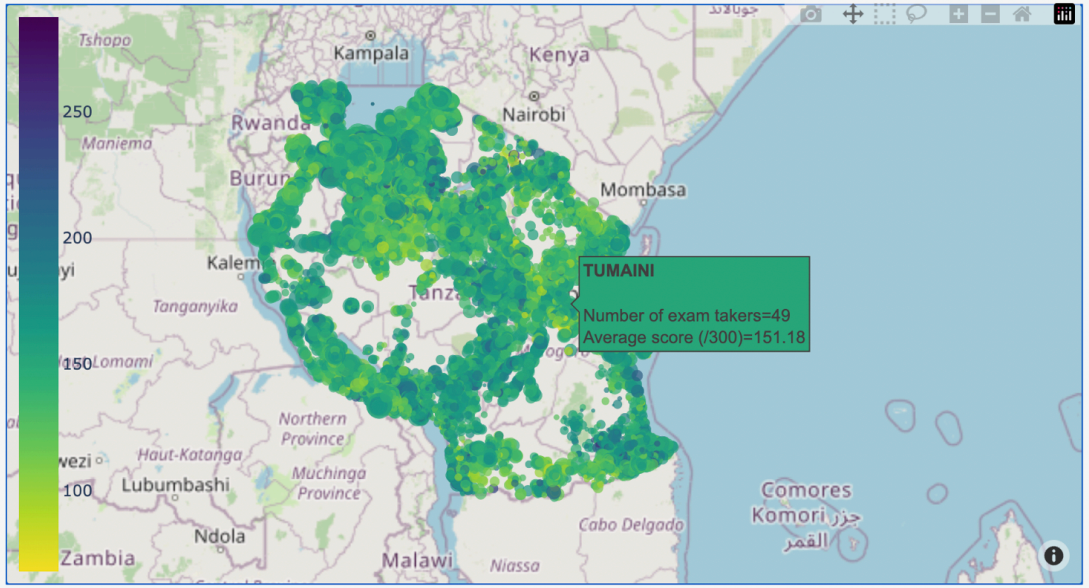
Schools (16361)

ALL SCHOOLS

Average score

This measures the average of the total scores of each student (per-school) in the national examination consisting of six subjects: Swahili, English, Social Studies, Mathematics, Science, Civics (50 marks each). Region data uses the unweighted mean average of all of its schools.

- **Unit:** Marks out of 300 total
- **Source:** NECTA PSLE Results (Published annually. Accessed 5 Aug, 2023.)



TUMAINI - PS1104104

Mkundi Ward, Morogoro Municipality, Morogoro Region
Type: Government, Total students: 445

PSLE Results Page

Number of exam takers	49
Number passed	40
Percent passed (A-C)	81.63
Average score (/300)	151.18
Average score grade	C
National quintile	middle

Resources Data

Pupil-to-Teacher Ratio	40.45
Pupil-to-Book Ratio Std7	3.58
Book-to-Pupil Ratio Std7	0.28
CG per student (TZS)	4,199
Ages mean (approx.)	9.26

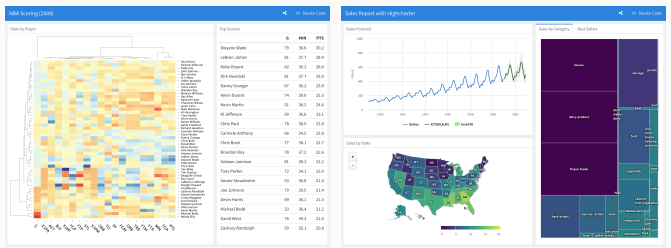
Geospatial Features

Pop. 3km radius	913
d to closest school (km)	3.85
d to council HQ (km)	17.3

`Source` Lonny Chen, NECTA PSLE Dashboard 2022 | https://bit.ly/psle2022mvp

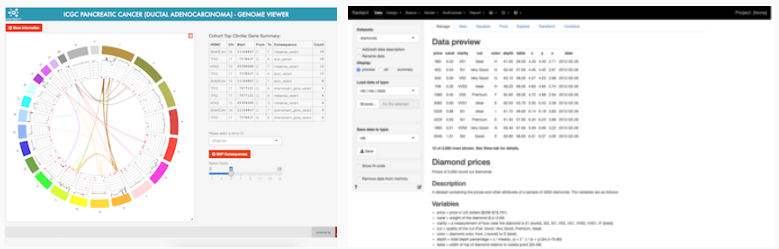
`flexdashboard` package

- Overview here.
- Good for easy dashboard building
- Just a document that looks like a dashboard
- Can be compiled into a static file (just like regular Markdown)
- Can only run interactive code client-fleside (in embedded JavaScript)
- Shiny and `htmlwidgets` (`leaflet`, `plotly`, `highcharter`, etc.) can be integrated (with all the up- and downsides)



`shiny` package

- Overview here.
- More complex to program, but the best option for complex apps.
- Can implement any layout.
- Needs a server behind it to execute R code on user input.
- Can run interactive code either by processing serverside (in R) or clientside (in embedded JavaScript).
- The shinydashboard package provides another way to create dashboards with Shiny.



Functionality

- Use simple R Markdown to build a dashboard.

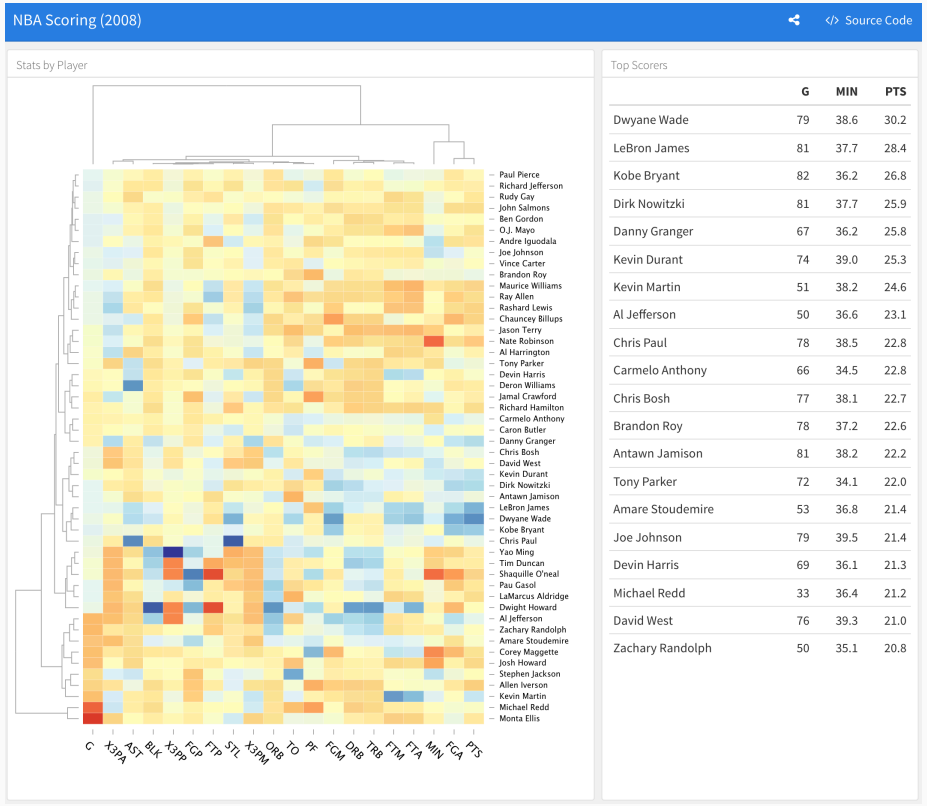
```
---
title: "NBA Scoring (2008)"
output:
  flexdashboard::flex_dashboard:
    orientation: rows
    social: menu
    source_code: embed
---

```{r setup, include=FALSE}
library(knitr)
library(d3heatmap)
library(flexdashboard)

url <- "http://datasets.flowingdata.com/ppg2008.csv"
nba_players <- read.csv(url, row.names = 1)
```

Stats by Player {data-width=650}

```{r}
d3heatmap(nba_players, scale = "column")
```

Top Scorers {data-width=350}

```{r}
knitr::kable(nba_players[1:20,c("G", "MIN", "PTS")])
```
```



	G	MIN	PTS
Dwyane Wade	79	38.6	30.2
LeBron James	81	37.7	28.4
Kobe Bryant	82	36.2	26.8
Dirk Nowitzki	81	37.7	25.9
Danny Granger	67	36.2	25.8
Kevin Durant	74	39.0	25.3
Kevin Martin	51	38.2	24.6
Al Jefferson	50	36.6	23.1
Chris Paul	78	38.5	22.8
Carmelo Anthony	66	34.5	22.8
Chris Bosh	77	38.1	22.7
Brandon Roy	78	37.2	22.6
Antawn Jamison	81	38.2	22.2
Tony Parker	72	34.1	22.0
Amare Stoudemire	53	36.8	21.4
Joe Johnson	79	39.5	21.4
Devin Harris	69	36.1	21.3
Michael Redd	33	36.4	21.2
David West	76	39.3	21.0
Zachary Randolph	50	35.1	20.8

Source: jjallaire

Functionality

- Use simple R Markdown to build a dashboard.
- Arrange panels as blocks with flexible syntax.

Layout by Column

By default, level 2 markdown headers (`----------`) within dashboards define columns, with individual charts stacked vertically within each column. Here's the definition of a two column dashboard with one chart on the left and two on the right:

````
---
title: "Column Orientation"
output: flexdashboard::flex_dashboard
---

Column
-------------------------------------

### Chart 1

```{r}
```

Column
-------------------------------------

### Chart 2

```{r}
```

### Chart 3

```{r}
```
````

Functionality

- Use simple R Markdown to build a dashboard.
- Arrange panels as blocks with flexible syntax.

SCROLLING LAYOUT

By default flexdashboard charts are laid out to automatically fill the height of the browser. This works well for a small number of vertically stacked charts, however if you have lots of charts you'll probably want to scroll rather than fit them all onto the page. You can control this behavior using the `vertical_layout` option. Specify `fill` to vertically re-size charts so they completely fill the page and `scroll` to layout charts at their natural height, scrolling the page if necessary.

For example, the following layout includes 3 charts and requests that the page scroll as necessary to accommodate their natural height:

````
---
title: "Chart Stack (Scrolling)"
output:
  flexdashboard::flex_dashboard:
    vertical_layout: scroll
---

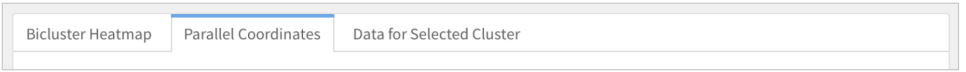
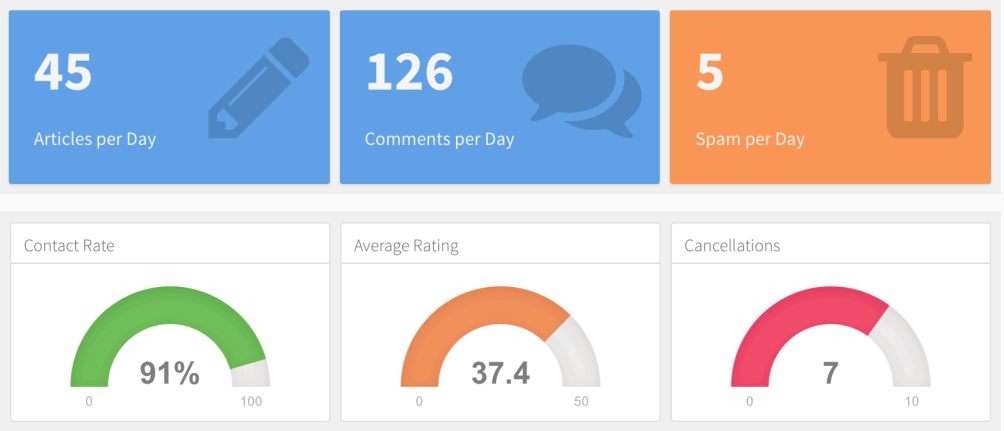
### Chart 1

```{r}
```

### Chart 2

```{r}
```

### Chart 3

```{r}
```
````

Functionality

- Use simple R Markdown to build a dashboard.
- Arrange panels as blocks with flexible syntax.

TABSETS

If you have several components you'd like to display within a row or column then rather than attempting to fit them all on screen at the same time you can lay them out as a tabset. This is especially appropriate when one component is primary (i.e. should be seen by all readers) and the others provide secondary information that might be of interest to only some readers.



In many cases tabsets are a better solution than `vertical_layout: scroll` for displaying large numbers of components since they are so straightforward to navigate.

To layout a row or column as a tabset you simply add the `{.tabset}` attribute to the section heading. For example, the following code lays out the second column in tabset:

```
---
title: "Tabset Column"
output: flexdashboard::flex_dashboard
---

Column
-------------------------------------

### Chart 1

```{r}
```

Column {.tabset}

Chart 2

```{r}
```

Chart 3

```{r}
```
```

Functionality

- Use simple R Markdown to build a dashboard.
- Arrange panels as blocks with flexible syntax.
- Add elements like gauges and value boxes.



Dashboards with `flexdashboard`

Functionality

- Use simple R Markdown to build a dashboard.
- Arrange panels as blocks with flexible syntax.
- Add elements like gauges and value boxes.
- Couple it with `shiny`.

````
---
title: "Old Faithful Eruptions"
output: flexdashboard::flex_dashboard
runtime: shiny
---

```{r global, include=FALSE}
# load data in 'global' chunk so it can be shared by all users of the dashboard
library(datasets)
data(faithful)
```

Column {.sidebar}
-----------------------------------------------------------------------

Waiting time between eruptions and the duration of the eruption for the
Old Faithful geyser in Yellowstone National Park, Wyoming, USA.

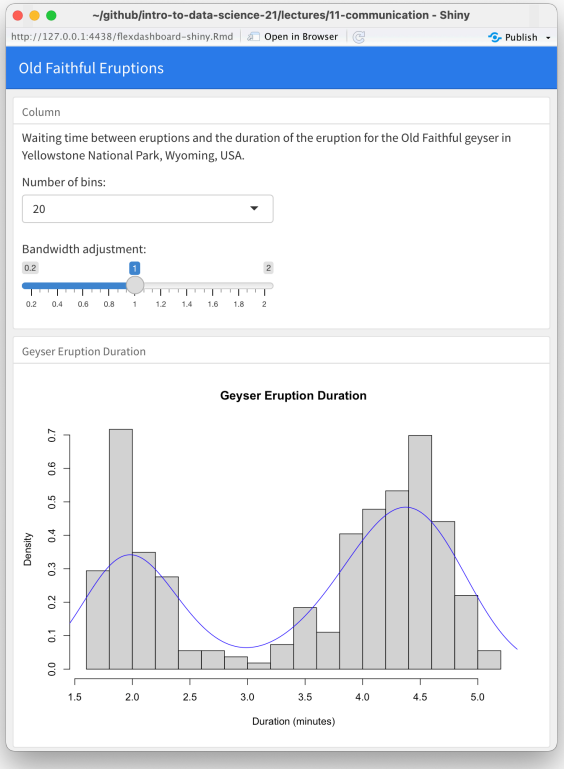
```{r}
selectInput("n_breaks", label = "Number of bins:",
            choices = c(10, 20, 35, 50), selected = 20)

sliderInput("bw_adjust", label = "Bandwidth adjustment:",
            min = 0.2, max = 2, value = 1, step = 0.2)
```

Column
-----------------------------------------------------------------------

### Geyser Eruption Duration

```{r}
renderPlot({
  hist(faithful$eruptions, probability = TRUE, breaks = as.numeric(input$n_breaks),
       xlab = "Duration (minutes)", main = "Geyser Eruption Duration")

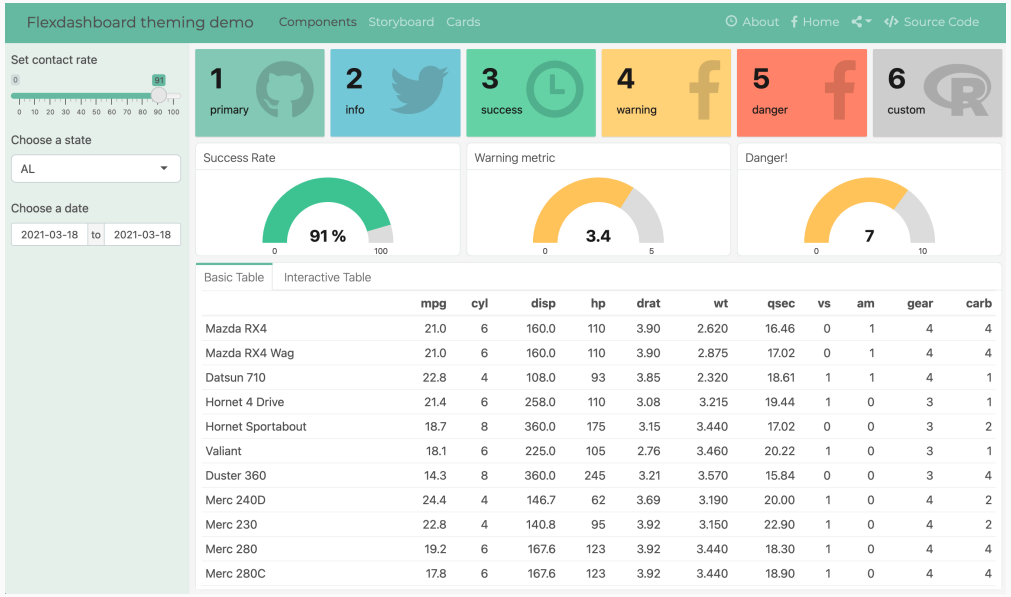
  dens <- density(faithful$eruptions, adjust = input$bw_adjust)
  lines(dens, col = "blue")
})
```
````



Dashboards with `flexdashboard`

Functionality

- Use simple R Markdown to build a dashboard.
- Arrange panels as blocks with flexible syntax.
- Add elements like gauges and value boxes.
- Couple it with `shiny`.
- Customize themes.

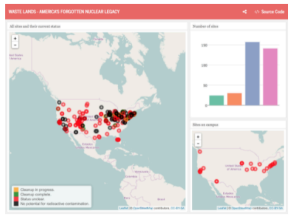


Functionality

- Use simple R Markdown to build a dashboard.
- Arrange panels as blocks with flexible syntax.
- Add elements like gauges and value boxes.
- Couple it with `shiny`.
- Customize themes.
- Explore more examples here.

MetricsGraphics: Tor Project

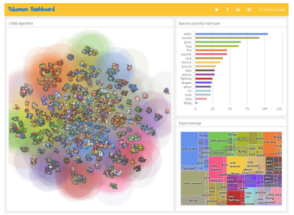
Shiny: kmeans clustering



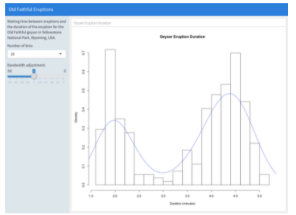
leaflet: nuclear waste sites

Shiny: biclust example

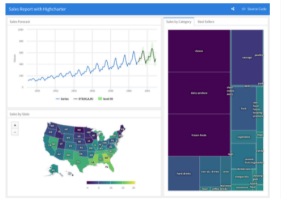
ggplotly: various examples



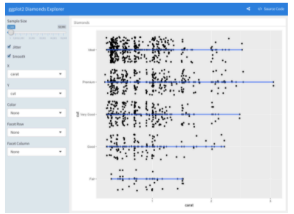
Pokemon characters with highcharter



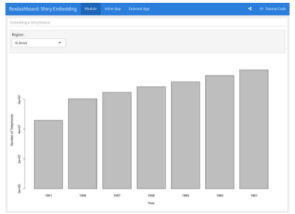
Shiny: Old faithful eruptions



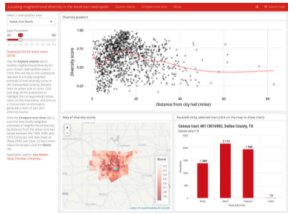
Sales report with highcharter



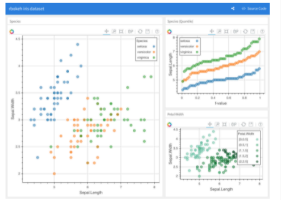
Shiny: Diamonds explorer



Shiny: Embedding



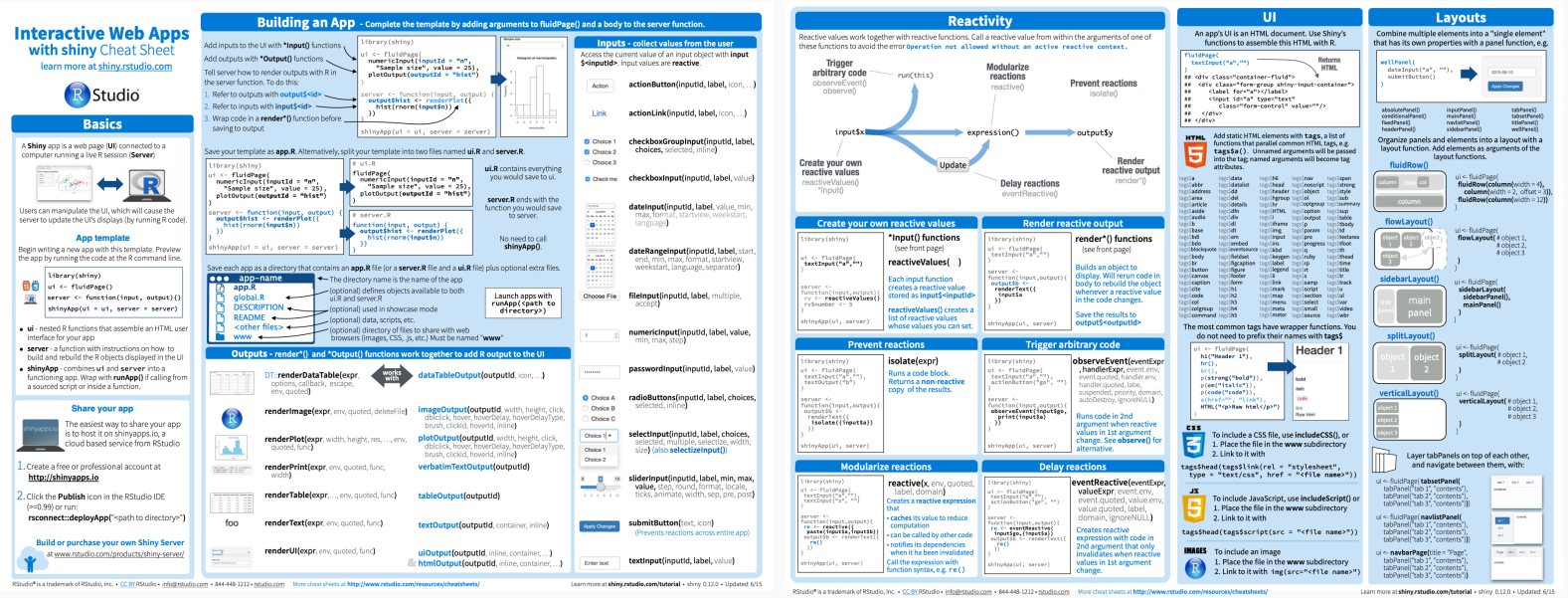
Shiny: Neighborhood diversity (Source)



rbokeh: iris dataset

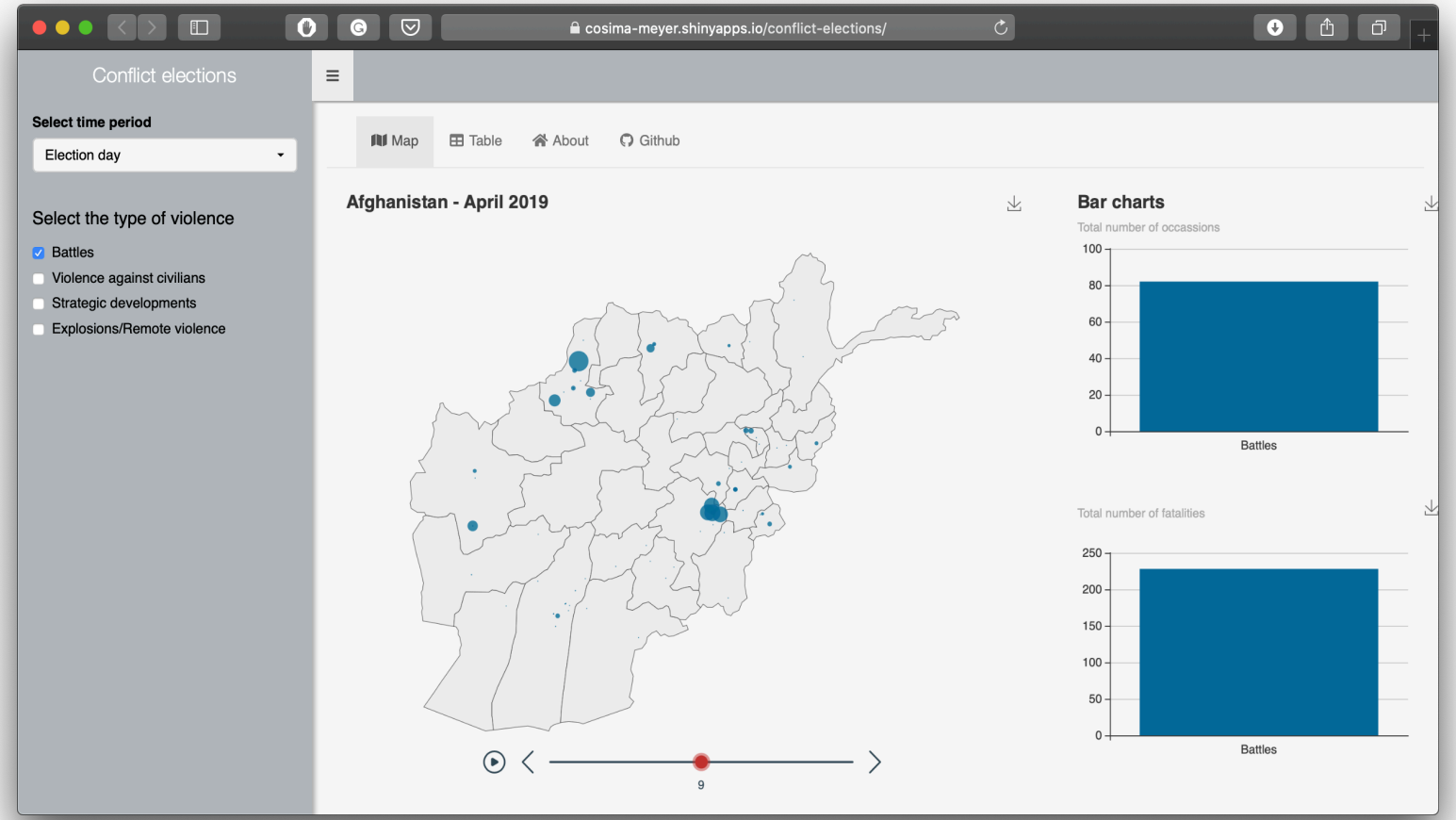
Functionality

- Shiny's functionality is too complex and rich to introduce it on a couple of slides. Wait for the labs!
- It certainly can do much more than dashboards.
- Think of it as a tool to create **web apps** that allow interaction with raw and cooked data.



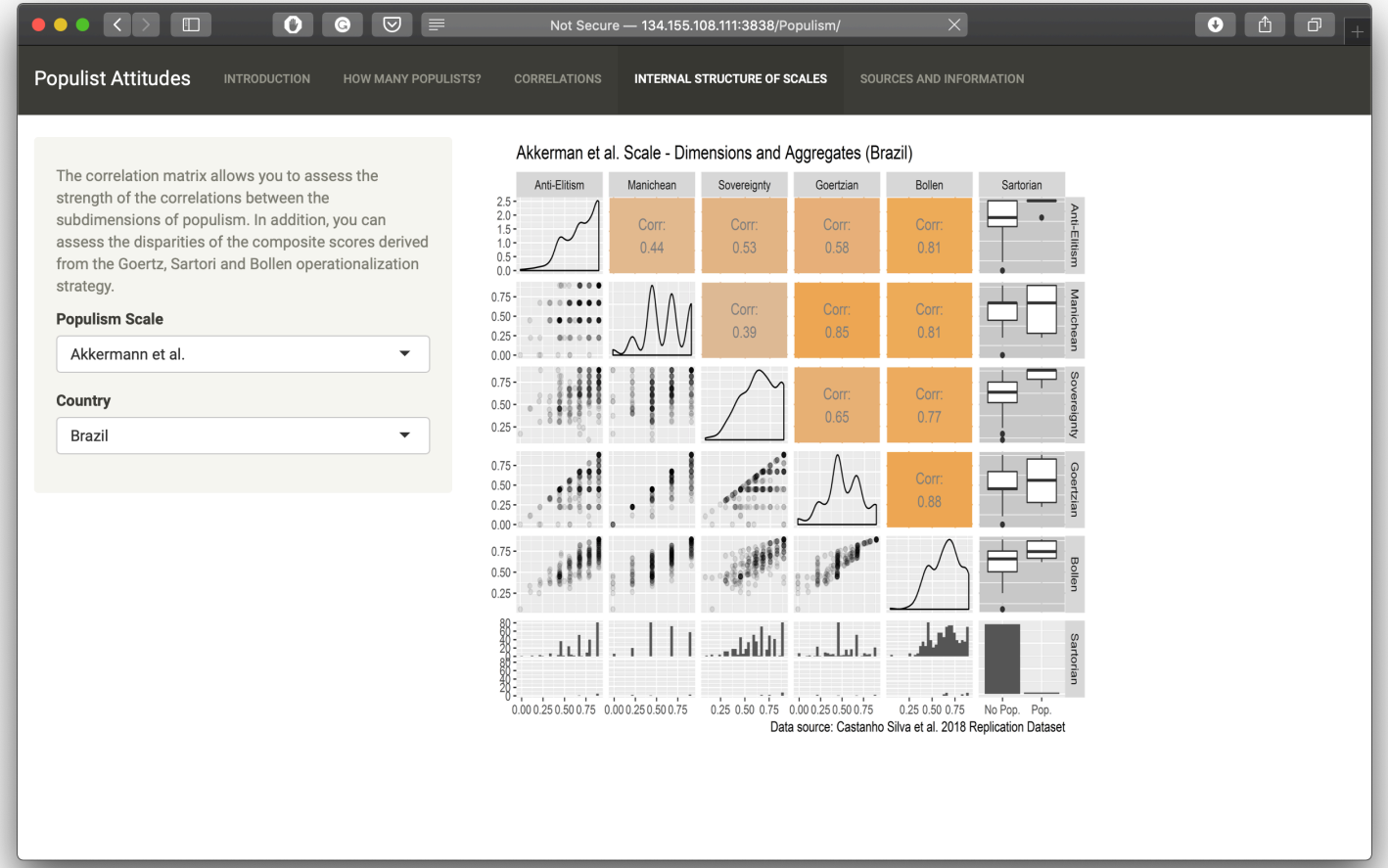
Example applications

- Data explorer



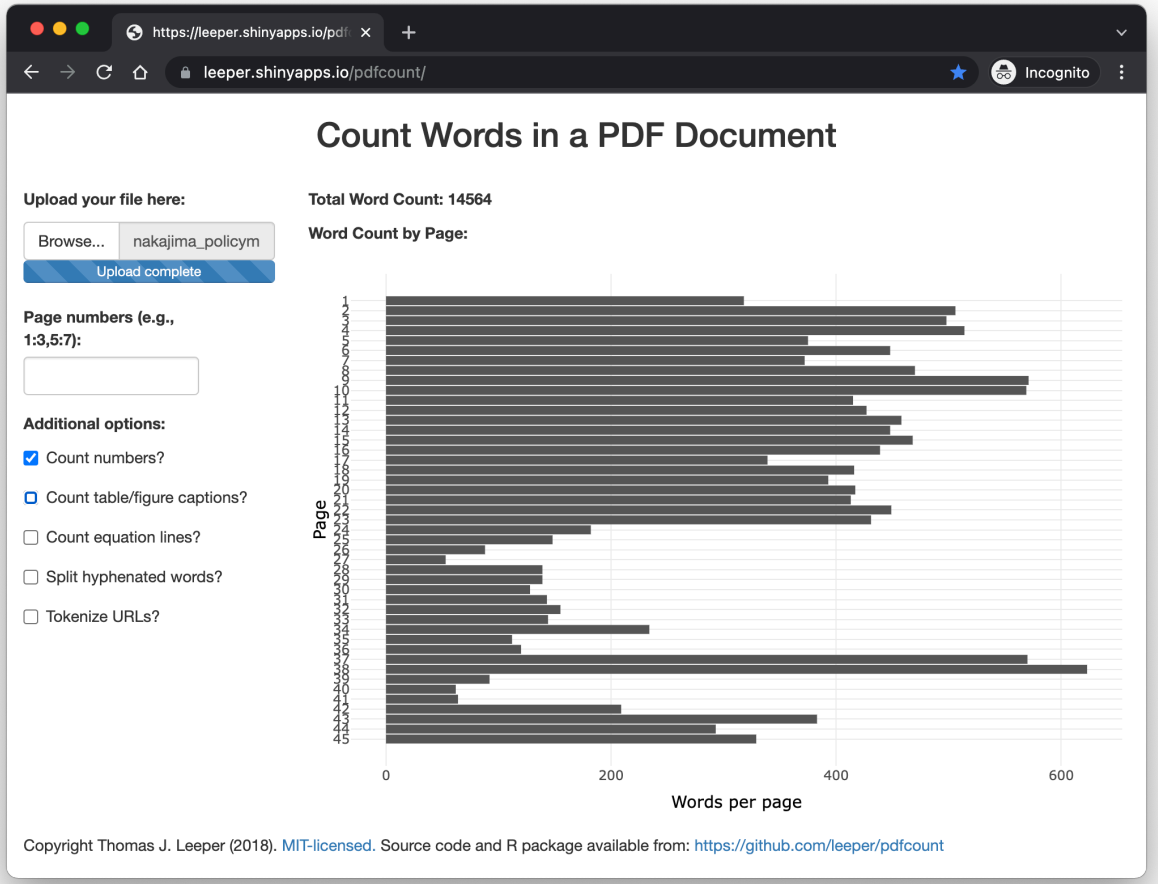
Example applications

- Data explorer
- Interactive appendix



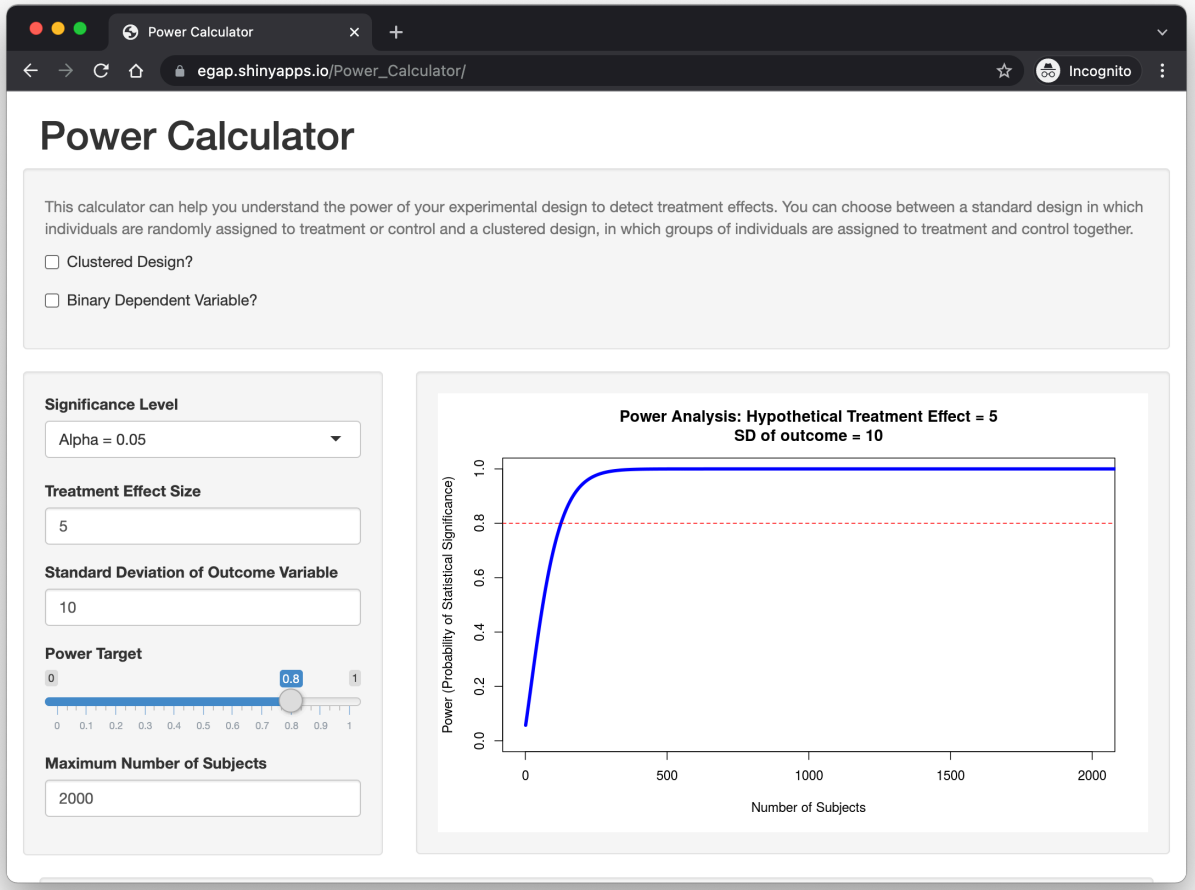
Example applications

- Data explorer
- Interactive appendix
- Workflow apps



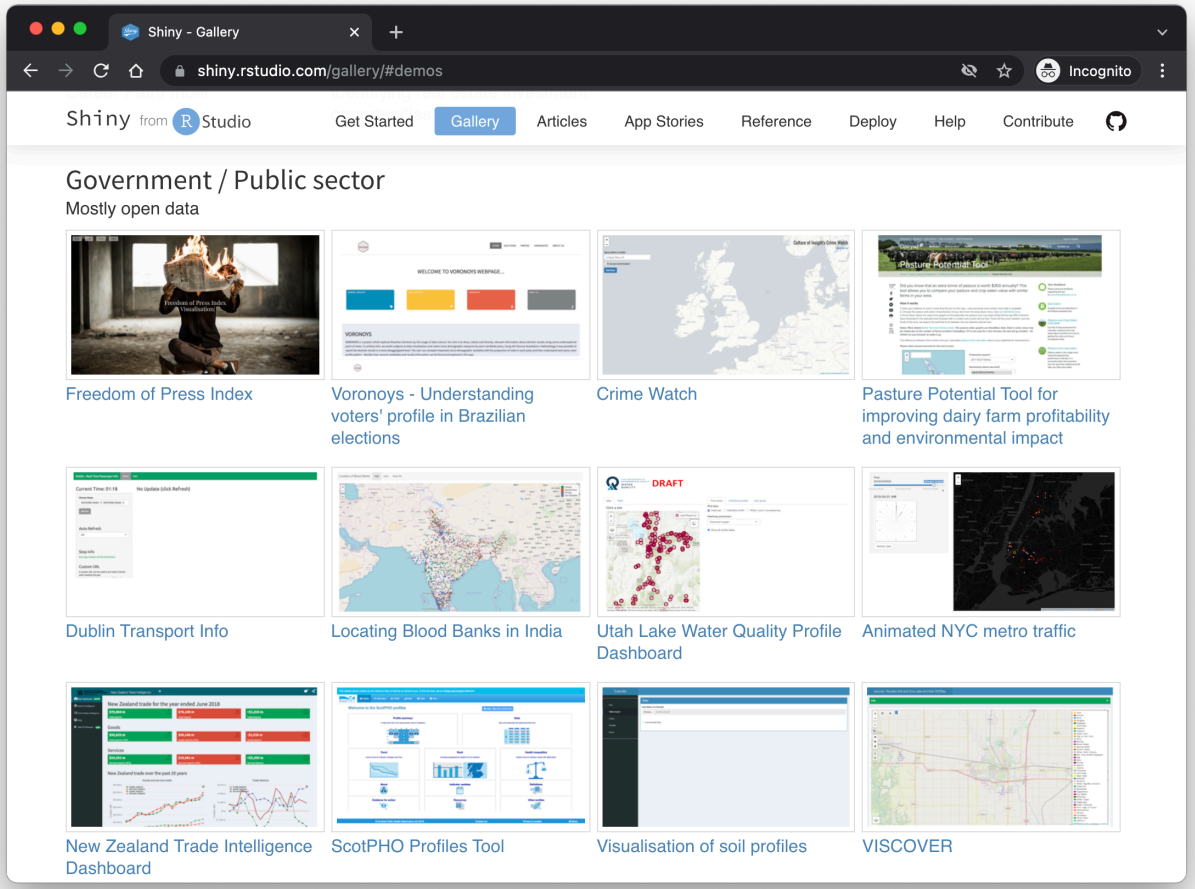
Example applications

- Data explorer
- Interactive appendix
- Workflow apps
- Learning tools



Example applications

- Data explorer
- Interactive appendix
- Workflow apps
- Learning tools
- ... and so much more!



More Shiny resources

Online resources

- Shiny official website
- Shiny official tutorial
- Shiny cheatsheet
- Mastering Shiny, book by Hadley Wickham
- Many useful articles about different topics
- Publishing own Shiny apps for free with shinapps.io
- Hosting your Shiny app on your own server
- Debugging Shiny

An overview of Shiny extensions

- awesome-shiny-extensions

Some highlights

- shinythemes: Altering the overall appearance of Shiny apps
- shinyjs: Enrich apps with JavaScript operations
- leaflet: Interactive maps
- ggvis: Similar to ggplot2 but with focus on web and interaction
- shinydashboard: Tools to create visual dashboards

Communicating as a data scientist

1. **Clarity and understanding**
   - Data science involves complex concepts.
   - Clear and concise explanations are necessary for non-technical stakeholders.
   - Encourages understanding to eventually make informed decisions based on data findings.
2. **Cross-organiational collaboration**
   - Data scientists work with stakeholders from various departments.
   - Understanding the organizational context is essential.
   - Findings should be communicated to show their impact on different parts of the organization.
3. **Adaptation to changes**
   - Staying updated and communicating changes effectively is vital for organizations.
   - Helps in keeping the team or organization aligned with the latest advancements.
   - Data science is rapidly evolving with new techniques and technologies.
4. **Articulating importance**
   - Data science projects can improve user experience, drive innovation, and optimize operations.
   - Clearly articulating the value of these projects is key to securing investment.
   - Essential for gaining support and resources from the team or organization.

Can we interchange data science with policy practitioner?

Why communication matters to us

Communication is central out the professional life as policy practitioners.

Reporting is an activity that we take on constantly, whether for colleagues, superiors, or the general public, **we create insights**.

There are ways in which we can encourage their use and support potential consumers

Identifying reporting requirements

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Communication plan

- Develop a plan that **outlines the strategies** which will be used to communicate the results of your monitoring and/or evaluation.

Reporting needs analysis

- Engage with the intended audience to **determine their reporting needs**.

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for individual learning (Traditionally)

Medium	Features
Short communications: Memos and Email, Postcards	- Can update audiences about ongoing evaluation activities and disseminate findings and recommendations in a timely manner - Can convey a limited amount of information in a confined length and format - Can be used to recap or follow-up on decisions or points made during a telephone or face-to-face meeting - Can be cost-effective means for communicating and reporting with a broad range of audiences and stakeholders - Are usually reserved for internal communications - May not be a confidential means of communication

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for individual learning (Traditionally)

Medium	Features
Interim/Progress Reports	- Are often used to present the results of partial evaluation findings - Are typically short reports but can be of considerable size in the case of multiyear, multisite, and/or very comprehensive evaluations - Are usually produced prior to a more comprehensive report - Should emphasize the provisional nature of interim reports to clients and stakeholders

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for individual learning (Traditionally)

Medium	Features
Final Reports	- Are the most commonly used form of reporting - Often uses a social science research format that includes the study's rationale, guiding questions, methodology, results, and recommendations - Provide an opportunity to portray the program in a more holistic and comprehensive manner - Should be designed using text, graphics, layout, and clear writing to increase accessibility and readability - Provide important archival documentation often for accountability purposes

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for individual learning (Traditionally)

Medium	Features
Executive Summaries	- Are written to make vital information more accessible - Are typically 1-5 pages in length - Briefly describe the evaluation's purpose, key questions, design, data collection and analysis methods - Emphasize the evaluation's findings and recommendations - Can be used separately from the final report

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for individual learning (Traditionally)

Medium	Features
Newsletters, Bulletins, Briefs, Brochures	- Can be disseminated to publicize upcoming evaluation activities - Can be distributed after evaluation activities to reinforce emergent findings or actions taken on the evaluation's findings - Provide opportunities to reach broad groups of people - Allow for presentation of text and graphics - Are typically 1-4 pages in length - Provides contact information

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for individual learning (Traditionally)

Medium	Features
Newsmedia	- Can disseminate evaluation information to a broad range of stakeholders and audiences - Can facilitate conversations among outside groups affected by or interested in the evaluation results - Can use mass media (newspapers, TV, radio) or specialized media (trade or interest group newsletters, journals, magazines) - Require careful thought and planning

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for individual learning (Traditionally)

Media	Features
Website and social media communications	- Make evaluation information available to a broad national and international audience at minimal cost - Provide an opportunity to stage the presentation of information into sections, accessible to users through links that the user selects - Provide useful information to evaluators who may be conducting evaluations on a similar topic or program - Increase potential reach of evaluation

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for group learning (Traditionally)

Media	Features
Verbal Presentations	- Are particularly useful for conveying information on complex, specialized topics - Should be focused on a few selected topics - Are most effective when audience members have an opportunity to interact with the information being presented - Should be accompanied with visuals such as overhead transparencies, PowerPoint slides, handouts, and flip charts

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for group learning (Traditionally)

Media	Features
Video & Computer Generated Presentations	- Are useful for conveying information about new or innovative programs - Are useful for disseminating findings to broader audiences than those directly involved with a program - Are effective for providing evaluation findings to groups whose time is limited and/or who might not be accustomed to reading evaluation reports - Can be used in multisite evaluations to depict events and activities at different sites - Are useful for documenting evaluation processes

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for group learning (Traditionally)

Media	Features
Posters and Poster Sessions	- Provide quick, visually oriented, easy to read information - Typically include photographs, diagrams, graphs, tables, charts, drawings, and text on poster-size boards - Provide a focused message with a clear purpose - Can stand alone as a visual display

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for group learning (Traditionally)

Media	Features
Working Sessions	- Identify concerns about the evaluation; establish buy-in - Are useful for making decisions regarding the evaluation's implementation and/or use of the findings - Can be used to obtain input about evaluation design, procedures - May be effective for engaging stakeholders and audiences in interpreting findings and developing recommendations/action plans

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for group learning (Traditionally)

Media	Features
Synchronous Electronic Communications	- Allow participants to exchange messages in real time through an electronically networked environment without being in each other's physical presence - Are useful when face-to-face meetings are impractical or impossible - Can lead to more timely dissemination of findings and subsequent decision-making - Provide opportunities to include stakeholders who might not otherwise be able or available to travel to a meeting - Provide opportunities for interaction and collaboration

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for group learning (Traditionally)

Media	Features
Personal Discussions	- May either be planned or impromptu, and may be initiated by the evaluator or by someone else - Occur face to face or over the telephone - Are one of the most powerful forms of communication; they facilitate insight, understanding, and new knowledge

`Source`: (Torres et al. 1996)

Developing reporting media

Ask yourselves: **What timeframe, format, and type of learning is required for reporting?**

Media for group learning (Traditionally)

Media	Features
Photography	- Can represent the realities of program participants; conveys various perspectives - Can illustrate the activities of a single participant in a program - Can be used to count, measure, compare, qualify, or track artifacts or information that can be captured visually - Is particularly useful when language or cultural barriers may inhibit participants' ability to verbally express their opinions, and/or easily assimilate information in written reports - May stimulate audiences' participation in interpreting important events and experiences, and enables them to use the findings

`Source`: (Torres et al. 1996)

Ensuring accessibility

Ask yourselves: **How can this information be easy to access and use for different users?**

- **Use appropriate language**: Make sure the language of a report is clear, concise and allows accessibility for all consumers. **AVOID UNNECESSARY JARGON**.
- **Apply graphic design principles**
- **Use simplified report layouts**: Three different ways of simplifying the report layouts are to eliminate chartjunk, emphasise headings as summary statements, and use descriptive subtitles.
- In evaluation reports, think of the **One-Three-Twenty-Five (1:3:25) principle**: Present findings in a logical and consistent manner by allowing *roughly* for a 1 page outline, a 3 page executive summary and 25 pages to present the findings and methodology.
- **Beware of the inclusiviy of your medium**: Think of users with auditory or visual disabilities.

How should we communicate today?

Hertie School

Policy documents in Georgia

Georgia's documents

Hertie School

1 **GENERAL FEATURES**
(purpose, period of time, structure)

2 **LEVEL OF IMPACT**
(policy documents hierarchy)

Georgia's documents

Hertie School

1 **GENERAL FEATURES**
(purpose, period of time, structure)

2 **LEVEL OF IMPACT**
(policy documents hierarchy)

According to international practice, the government policy can be represented in different types of policy documents, such as: Programme, Concept, Strategy, Action Plan, Policy Brief, White Paper and Green Paper, laws or bylaws, speeches of public officials, etc. However, in policy planning and coordination system of Georgia 3 types of documents are designated: a concept (document), a strategy (document) and an action plan, whereas the development of other documents mentioned above is governed by other legislative and methodological regulations or may be subject to future regulation. Figure 1 shows a **mandatory structure** of all the three types of policy documents, which must be followed content-wise.

Policy Document

A document of strategic or operational nature that defines a national or sector policy and outlines solutions to problems and ways of developing of the area.

Georgia's documents

Hertie School

1 **GENERAL FEATURES**
(purpose, period of time, structure)

2 **LEVEL OF IMPACT**
(policy documents hierarchy)

Concept (document)

National or sector policy document of general nature that defines the need of strategy development, a vision, basic principles and priorities.

CONCEPT - in addition to the concept of the Parliament of Georgia, which defines sectoral priorities of policy development in a field, the government may also adopt a concept document on the development of and solution to a problem in an area or an institution. Preparation of such a document is not mandatory and the government or a relevant coordination body determines the need for it. A concept is adopted to save resources by avoiding the development of a policy that is not needed or does not comply with the government priorities and objectives, or when various institutions fail to agree on principle issues for the development of a particular area.

Georgia's documents

Hertie School

1 **GENERAL FEATURES**
(purpose, period of time, structure)

2 **LEVEL OF IMPACT**
(policy documents hierarchy)

STRATEGY - a government policy in a particular field is mainly expressed in a strategy document. A strategy, if any, must be based on a concept document adopted by the Parliament of Georgia or developed by the government and the **goals/objectives** specified in the strategy must respond to the vision, principles and sector priorities of the concept (if any). However, in the majority of cases, a strategy is developed without a concept document. The **introductory part** of the strategy must substantiate the need of its development. A strategy document formulates the government policy for tackling a problem and describes specific mechanisms and intended results that must be achieved through the implementation of the strategy. A strategy must be based on priority directions identified in national policy documents and ensures the continuation of the implementation even after the change of government.

Strategy (document)

A policy document that identifies priorities, goals and objectives for addressing identified national or sectoral problems, outlines approaches to the implementation of set goals/objectives and determines performance indicators to track progress.

1 GENERAL FEATURES
(purpose, period of time, structure)

2 LEVEL OF IMPACT
(policy documents hierarchy)

Action Plan

A policy document that defines specific activities for achieving sectoral priorities, goals and objectives, their output indicators, responsible and partner agencies, deadlines, budget and source of financing.

ACTION PLAN - operational level policy document, designed to outline specific activities to be carried out to fulfill the goals and objectives set in the strategy and achieve the results defined in the strategy.

An action plan must, at least, contain the information provided in Figure 1. Other information may be added to it, if so decided.

Duration of action plan ranges between minimum one and maximum three years. However, in exceptional cases, it is acceptable to adopt six-month or quarterly action plans too.

In some cases, an action plan may be adopted as an independent document, without a corresponding strategy **(sector action plan)**. Such action plan must include a narrative section offering a brief situation analysis, defining objectives, as well as implementation, monitoring and evaluation mechanisms.

Georgia's documents

Hertie School

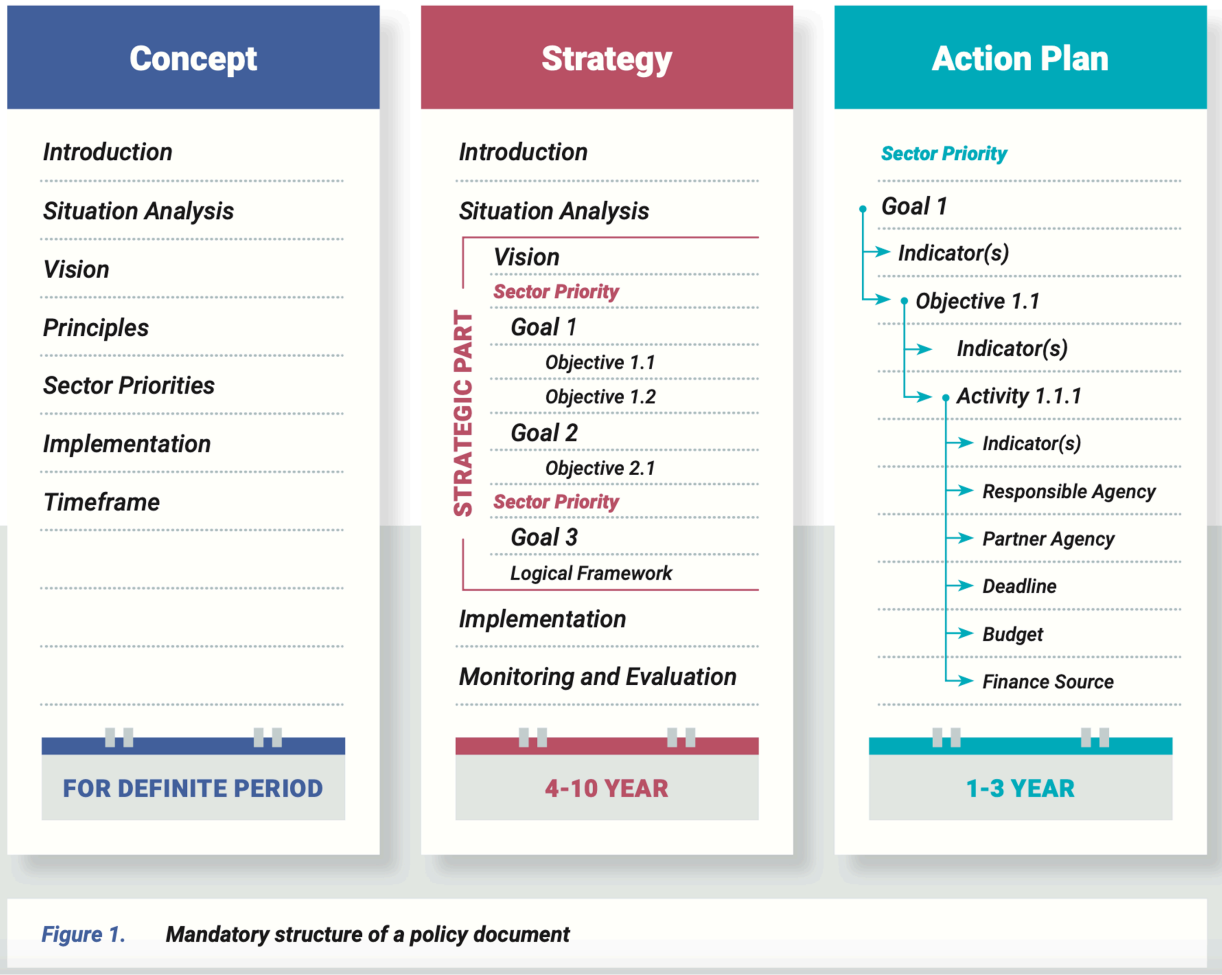


Figure 1. **Mandatory structure of a policy document**

Georgia's documents

Hertie School

1 NATIONAL POLICY DOCUMENTS;

2 SECTOR POLICY DOCUMENTS;

3 INSTITUTIONAL POLICY DOCUMENTS.

#	LEVEL	NAME OF POLICY DOCUMENT	COORDINATING AGENCY	ADOPTION
1	NATIONAL	National development strategy	The Administration of the Government	Government
		Government program	Winning political party (parties) / administration of government /ministries/LEPLs	Parliament
		Basic Data and Directions (BDD) - MTEF → *State budget*	Ministry of Finance	Parliament
		Annual Government Work Plan	The Administration of the Government	Government
		International Agreements / Commitments	According to legislation	According to legislation
2	SECTOR	Multisector and sector **strategies** → *Relevant action plans*	The Administration of the Government, Ministries, State Minister's Office or LEPLs accountable to the government	Government
		Multisector and sector **action plans**	The Administration of the Government, Ministries, State Minister's Office or LEPLs accountable to the government	Government
3	INSTITUTIONAL	Institutional development strategies → *Relevant action plans*	Public institutions	Government or ministry or institution
		Mid-term action plans	Ministries	Ministry

Figure 2. ***Hierarchy of policy documents.***

Questions?